

**Tiedon visualisoinnin prosessi ja sen soveltaminen
ammattikorkeakoulun opintoasiainhallintoon**

Katariina Lehto

Tampereen yliopisto
Informaatiotieteiden yksikkö
Tietojenkäsittelyoppi
Pro gradu -tutkielma
Ohjaajat: Timo Niemi, Marko Junkkari
Marraskuu 2015

Tampereen yliopisto
Informaatiotieteiden yksikkö
Tietojenkäsittelyoppi
Katariina Lehto: Tiedon visualisoinnin prosessi ja sen soveltaminen
ammattikorkeakoulun opintoasiainhallintoon
Pro gradu -tutkielma, 73 sivua
Marraskuu 2015

Tiivistelmä

Tässä tutkielmassa tarkastellaan tiedon visualisointia prosessinäkökulmasta. Tarkoituksena on luoda kokonaiskuva visualisointiprosessin eri vaiheista, niiden suhteista toisiinsa sekä eri tieteenaloilla tehtyjen tutkimuksien tulosten merkityksestä onnistuneen visualisoinnin syntymiseen. Tähän päämäärään pyritään tarkastelemalla aihepiiriä koskevaa eri alojen kirjallisuutta sekä tutustumalla tutkimuksissa esitettyihin visualisoinnin prosessimalleihin. Kirjallisuuden pohjalta tutkielmassa tehdään synteesi visualisointiprosessista, jonka pohjalta kehitetään lähestymistapa tiedon visualisointiin. Sen eri vaiheita käydään läpi vaihe vaiheelta käytännön sovellutusesimerkkien avulla.

Tutkielmassa visualisointiprosessia sovelletaan käyttäjälähtöisen suunnittelun periaatteita käyttäen ammattikorkeakoulun opintoasiainhallintoa kohdealueena. Prosessin tavoitteena on visualisoida ammattikorkeakoulujen yhtä keskeisintä tuloksellisuusmittaria eli 55 opintopistettä lukuvuodessa suorittaneita opiskelijoita. Erilaisilla visualisoinneilla pyritään tuottamaan uutta ymmärrystä ilmiöstä kohdealueen asiantuntijoille.

Avainsanat ja -sanonnat:

visualisointi, visualisointiprosessi, opintoasiainhallinto

Sisällys

1.	Johdanto.....	1
1.1.	Dataa, informaatiota vai tietoa.....	1
1.2.	Tiedon visualisoinnin määritelmä.....	3
1.3.	Tutkimuksen tavoitteet, menetelmät, rajaukset ja rakenne.....	5
2.	Tiedon visualisointi monitieteellisenä tutkimuskohteena.....	7
2.1.	Visualisoinnin merkitys ja tutkimus.....	7
2.2.	Näkötutkimus ja visuaalinen havainnointi.....	8
2.3.	HCI (Human Computer Interaction).....	11
3.	Tiedon visualisoinnin prosessi.....	13
3.1.	Visualisointiputki.....	13
3.2.	Tietämyksen muodostamisen prosessimalli.....	15
3.3.	Toryn ja Möllerin malli.....	16
3.4.	Van Wijkin prosessimalli.....	17
3.5.	Chittaron visualisointiprosessin tehtävälista.....	18
3.6.	Munznerin sisäkkäinen prosessimalli.....	19
3.7.	Yhtenäinen visualisointimalli.....	20
3.8.	Network Analysis and Visualization (NAV) –prosessimalli.....	22
3.9.	Visualisoinnin työkulkumalli.....	24
3.10.	Waren prosessimalli.....	25
3.11.	Malyanov et al. visualisointiprosessi.....	26
3.12.	Yhteenveto ja pohdintaa visualisointiprosesseista.....	27
4.	Tutkielman lähestymistapa visualisointiprosessiin.....	29
5.	Tavoitteen ja kohdealueen määrittely.....	31
5.1.	Tavoitteen määrittely.....	31
5.2.	Kohdealueen määrittely.....	32
5.2.1.	Käyttjäanalyysi.....	32
5.2.2.	Tehtäväanalyysi.....	35
5.3.	Menetelmiä.....	36
5.3.1.	Yhteistyömallit.....	36
5.3.2.	Tehtäväanalyysin menetelmiä.....	38
5.3.3.	Aikaiset prototyypit tavoitteenasettelussa.....	39
6.	Datan ja operaatioiden määrittely.....	43
6.1.	Datajoukon valinta.....	45
6.2.	Datatyypin määrittely.....	47
6.3.	Datan puhdistaminen (data cleaning).....	48
6.4.	Operaatioiden määrittely.....	50

7.	Visuaalisen ratkaisun määrittely ja toteutus	53
7.1.	Esitystavan valinta	53
7.2.	Dataa kuvaavien visuaalisten elementtien valinta	60
7.3.	Värien käyttö.....	61
7.4.	Selitteiden käyttö.....	62
7.5.	Operaatioiden toteutus	62
8.	Evaluointi.....	65
8.1.	Vaiheittainen evaluointi	65
8.2.	Kokonaisevaluointi	67
8.3.	Visualisoinnin evaluoinnin erityishaasteet.....	71
9.	Yhteenveto.....	72
	Viiteluettelo	74

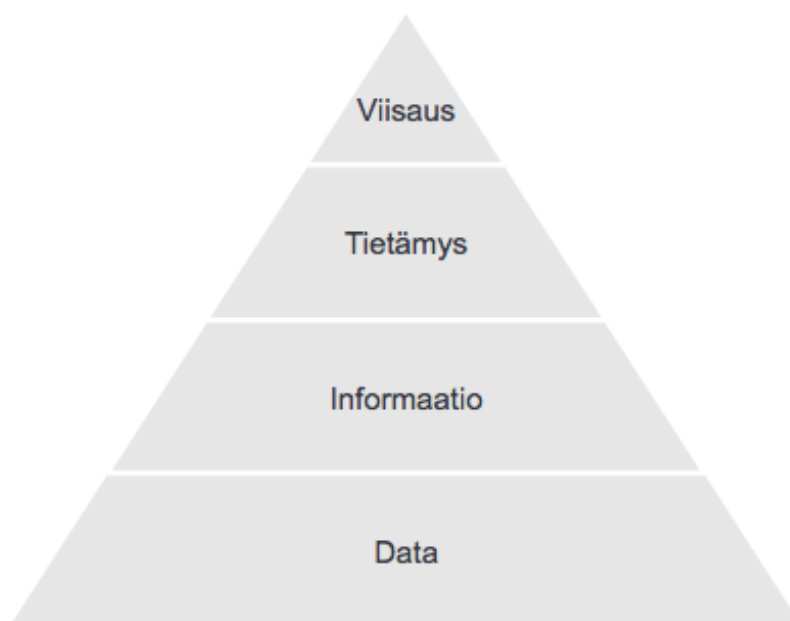
1. Johdanto

Tiedon visualisoinnin perimmäisenä tarkoituksena on luoda ihmisen havainnointikykyä hyödyntäviä visuaalisia esityksiä tietosisällöstä, jotta ihminen pystyisi tehokkaasti tulkitsemaan datan sisältöä ja hahmottamaan rakennetta. Visualisoinnilla voi löytää ja havainnollistaa muuten näkymättömiä suhteita erilaisten muuttujien välillä ja koostaa suurenkin datan sisältöä nopeasti hahmotettavissa olevaksi informaatioksi.

Johdannossa pureudutaan ensin datan, informaation ja tietämyksen käsitteiden keskinäisiin suhteisiin sekä määritellään, mitä tiedon visualisoinnilla tutkielmassa ymmärretään. Johdannon viimeisessä osassa esitellään tutkimuksen tavoitteet, valitut menetelmät, tehdyt rajaukset sekä yleisellä tasolla tutkielman rakenne.

1.1. Dataa, informaatiota vai tietoa

Datan, informaation ja tietämyksen käsitteiden keskinäistä suhdetta ja merkitystä on yritetty määrittellä ja kirjoittaa auki jo vuosikymmeniä, mutta siitä huolimatta niiden käytössä on edelleen huomattavaa ristiriitaisuutta ja päällekkäisyyttä. Määritelmiä on tehty useista eri näkökulmista tietojenkäsittelytieteessä ja tietotekniikassa sekä myös muilla tieteenaloilla kuten psykologiassa, johtamistieteissä ja epistemologiassa. Käsitteet ovat säilyttäneet monitulkintaisuutensa myös visualisointiyhteisössä, vaikka visualisoinnin tutkijat ovatkin niiden taksonomiaa yrittäneet selventää useita kertoja. [Chen et al., 2009]



Kuva 1. DIKW-pyramidi.

Yksi suosittu esitystapa eritellä ihmisen ymmärrystä on DIKW-pyramidi (data-information-knowledge-wisdom) (kuva 1), jonka hierarkiaan viittasi ensimmäisenä runoilija T.S. Eliot [1934]. Informaatiotieteiden kontekstissa siitä kirjoitti ensimmäisenä Harland Cleveland [1982]. Usein pyramidin alkuperäisinä keksijöinä erityisesti tietämyksenhallinnan tutkimuksissa kuitenkin mainitaan Russell Ackoff [1989] sekä Milan Zeleny [1987].

Ackoffin [1989] määritelmän mukaan data on vain joukko symboleja, jotka ilmaisevat asioiden, tapahtumien ja ympäristön ominaisuuksia ilman minkäänlaista tulkintaa eli ilman semanttista merkitystä. Kun käytettävissä olevaan dataan liitetään jokin konteksti ja merkitys, sitä voidaan kutsua informaatioksi [Ackoff, 1989]. Informaatio vastaa yleensä kysymyksiin kuka, mitä, missä, milloin tai montako [Rowley, 2007]. Vasta kun ihminen soveltaa tai sisäistää havainnoimaansa informaatiota, siitä muodostuu tietämystä, joka vasta kysymykseen miten [Chen et al., 2009]. Pyramidin kolmella alimmalla tasolla Zelenyn [1987] jaottelu on samankaltainen kuin Ackoffin, mutta määritelmät on muotoiltu hieman toisella tavalla. Zelenyn mukaan data ei tiedä mitään (know nothing), informaatio tietää mitä (know what) ja tietämys tietää miten (know how).

Tietojenkäsittelyn kontekstissa data, informaatio ja tietämys on määritelty olevan eri asioita kuin data, informaatio ja tietämys ihmiselle. Chen et al. [2009] määritelmän mukaan data on tietojenkäsittelyn menetelmin esitettyä todellisten tai simuloitujen olioiden malleja ja ominaisuuksia. Informaatio on jalostettua dataa, joka esittää tuloksia jostakin dataan kohdistuvasta tietojenkäsittelyprosessista, kuten tilastollisesta analyysistä tai merkitysten liittämistä dataan. Informaatio voi olla myös tallennettuja kuvauksia ihmisen mielessä syntyneistä merkityksistä. Tietämys määrittää informaationa, joka esittää tuloksia tietokoneella simuloituista kognitiivisista prosesseista, kuten havainnointi, oppiminen, assosiaatio tai päättely. Tietämys voi olla yhtä hyvin myös ihmisen toiminnan tuloksena syntynyttä, jos se on tallennettu tietojenkäsittelyn menetelmiä hyödyntäen.

Näistä näkökulmista tarkastellen suomen kielen käsite “tiedon visualisointi” on varsin monimerkityksinen ja jopa hieman harhaanjohtava. Tämä johtuu siitä, että sitä käytetään yleisesti kuvaamaan visualisointia useilla tiedon eri abstraktiotasoilla. Erityisesti raakadatan visualisoinnit, kuten sen jakaumiin ja monimuotoisuuteen liittyvät grafiikat, ovat DIKW-pyramidin perusteella luokiteltuna pikemminkin datan tai informaation visualisointeja. Onko kyseessä datan vai informaation visualisointi riippuu datalle tehdyn esitulkinnan ja -käsittelyn määrästä. Tällaiset visualisoinnit eivät todellisuudessa pyrikään visualisoimaan tietämystä, vaan sen sijaan ne pyrkivät auttamaan sen kehittymistä visualisoinnin kohderyhmän käyttäjissä. Visualisoinnin käyttäjä muodostaa assosiaatioita asioiden välille hyödyntäen sekä visualisoinnin antamaa informaatiota, että omaa aiempaa kohdealueen osaamistaan ja ymmärrystään. Staattiset selittävät visualisoinnit sen sijaan perustuvat usein vahvemmin aiempaan analyysiin ja ennalta tehtyihin päätelmiin

[Iliinsky & Steele, 2011]. Tällaiset visualisoinnit saattavat toisinaan välittää visualisoinnin suunnittelijan datasta ja informaatiosta ennalta hankkimaa tietämystä, jota halutaan esittää visuaalisin keinoin. Termillä tiedon visualisointi ja sen englanninkielisillä moninaisilla vastineilla käsitetään käytännössä kaikkea tätä.

Tässä tutkielmassa esimerkkinä sovellettava visualisointiprosessi asettuu parhaiten DIKW-pyramidin tasolle informaation visualisointi. Esimerkkivisualisoinnit perustuvat talletettuun dataan, jolle on annettu merkityksiä kohdealueen kontekstissa ja jota on käsitelty tilastollisin menetelmin (esim. summa, keskiarvo tai mediaani).

1.2. Tiedon visualisoinnin määritelmä

Tiedon visualisointi -käsitettä käytetään monessa eri merkityksessä myös muulla tavoin kuin pelkästään DIKW-pyramidin näkökulmasta. Sillä kuvaillaan yhtä lailla visualisointiin keskittyvää tieteellistä tutkimusta sekä visualisointiteknologioita ja -tekniikoita (esim. visualisoinnin menetelmiä ja työkaluja) kuin visualisointiprosessin visuaalista lopputuotettakin [van Wijk, 2006 I]. Näille visualisoinnin eri ulottuvuuksille ei ole olemassa muita erillisiä käsitteitä, joten myös tässä tutkielmassa visualisoinnin käsitettä käytetään asiayhteydestä riippuen kaikissa sen eri merkityksissä.

Tiedon visualisoinnille on englannin kielessä ainakin kolme toisistaan poikkeavaa käsitettä, joita käytetään kirjallisuudessa ja tutkimuksissa osin synonyymeina, osin päällekkäisinä ja toisinaan lähes erillisinä käsitteinä:

- data visualization,
- information visualization sekä
- scientific visualization [Ziemkiewicz & Kosara, 2007].

Alalla ei ole vielä saavutettu yksimielisyyttä näiden käsitteiden täsmällisistä määritelmistä ja keskinäisistä eroavaisuuksista, mutta niiden luominen olisi oleellista tulevien tutkimusten kannalta [Tory & Möller, 2004].

Selkeimmät eroavaisuudet on jossain määrin yhteisymmärryksessä kyetty löytämään tieteellisen visualisoinnin (scientific visualization) ja tiedon visualisoinnin (information visualization) käsitteiden välille. Tieteelliselle visualisoinnille on tyypillistä, että datalle löytyy jokin fyysinen vastine todellisuudesta ja dataa käsitellään tieteellisessä kontekstissa. Tiedon visualisointi sen sijaan keskittyy abstraktia todellisuutta kuvaavan datan visualisoimiseen muussa kuin tieteellisessä tarkoituksessa [Card et al., 1999]. Toisaalta kirjallisuudessa on myös kritisoitu edellä mainittua keinoitekoista rajanvetoa tieteellisen tiedon ja muun tiedon visualisoinnin välille, sekä pyritty löytämään toisenlaisia tapoja luokitella visualisointeja [Tory & Möller, 2002].

Tiedon visualisoinnilla on vielä vuonna 1972 ymmärretty visuaalisen mielikuvan muodostamista ajattelun avulla [Burchfield, 1972]. Nykyisin sillä tarkoitetaan yleisemmin graafista esitystä datasta tai tällaisen visuaalisen esityksen kehittämismenetelmiä. Käsitys visualisoinnista on ajan myötä ja tietojenkäsittelyn mahdollisuuksien lisääntyessä ja yleistyessä muuttunut ihmisen sisäisestä toiminnosta ulkoiseksi menetelmäksi. Ihminen voi kuitenkin hyödyntää sitä tehokkaasti päätöksenteon ja ymmärryksensä lisäämisen apuna. [Ware, 2012] Eräs tarkoitus on nopeuttaa ilmiön ymmärtämistä luomalla siihen korkealla abstraktiotasolla oleva visualisointi [Niemi, 2015].

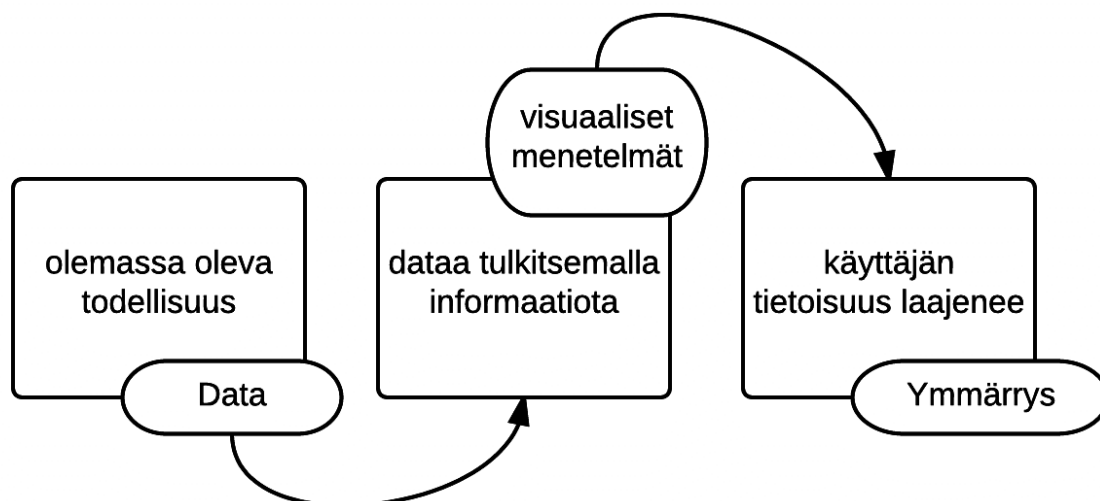
Vuoden 1995 Symposium of Information Visualization alkusanoissa [Gershon & Eick, 1995] tiedon visualisointi määritellään prosessina. Tässä prosessissa ilmiöön liittyvää dataa ja informaatiota, joille ei löydy luonnostaan konkreettista vastinetta todellisuudesta, muutetaan visuaaliseen muotoon. Tällöin käyttäjä voi sen avulla havainnoida ja ymmärtää ilmiötä.

Card et al. [1999] määritelmän mukaan visualisointi on datasta luotujen tietokonepohjaisten vuorovaikutteisten esitysten käyttöä tietoisuuden laajentamiseksi. Tämä näkemys rajaa käytännössä pois kaikki ei-vuorovaikutteiset, staattiset tiedon visualisoinnit (esim. painotuotteissa, kuten kirjoissa ja lehdissä käytetyt visualisoinnit). Niiden merkitys tiedon visualisoinnin tutkimuksessa ja erityisesti visualisoinnin historiassa on kuitenkin varsin olennainen. Card [2012] onkin myöhemmin laajentanut aiempaa määritelmää: Tiedon visualisointi on joukko teknologioita, jotka käyttävät visuaalista tietojenkäsittelyä laajentamaan ihmisen tietoisuutta olemassa olevasta informaatiosta.

Ben Shneiderman [2001] määrittelee tiedon visualisoinnin kompaktina graafisena esityksenä ja käyttöliittymänä, jolla voi käsitellä nopeasti suuria määriä tietoa. Toimivien visualisointien avulla käyttäjät voivat tehdä datasta löydöksiä, päätelmiä sekä havaita selittäviä tekijöitä. Havainnot voivat kohdistua esimerkiksi tietoalkioiden välisiin suhteisiin (esim. korrelaatioon, klustereihin, poikkeaviin arvoihin), tietojoukkoihin tai yksittäisiin tietoalkioihin.

GENE et al. [2005] määrittelevät tiedon visualisoinnin tekniikoiksi ja työkaluiksi, joilla dataa voidaan muuttaa sellaiseksi informaatioksi, joka lisää ilmiön ymmärrystä ja näkemystä. Tiedon visualisointi pyrkii esittämään datan sisältöä erilaisten visuaalisten primitiivien avulla niin, että abstraktista datasta pystytään muodostamaan graafinen esitys.

Kaikkien edellä esiteltyjen määritelmien perusajatus on erilaisista abstraktiotasoista ja yksityiskohdista huolimatta samankaltainen (kuva 2). Tiedon visualisoinnissa hyödynnetään visuaalisia menetelmiä olemassa olevien ilmiöiden esittämiseen niin, että ihmisellä on mahdollisuus havainnointikykyjensä avulla nopeasti laajentaa tietämystään kohdealueesta.



Kuva 2. Tiedon visualisoinnin määritelmien yhteiset piirteet.

Kirjallisuudessa esitetyt määritelmät perustuvat pääosin visualisoitavan datan olemukseen, tiedon visualisoinnin tarkoitukseen ja toteutustapaan sekä kohdealueeseen, jonka tarpeisiin visualisointia ollaan toteuttamassa.

1.3. Tutkimuksen tavoitteet, menetelmät, rajaukset ja rakenne

Tutkielman tarkoituksena on luoda kokonaiskuva tiedon visualisoinnin prosessin eri vaiheista, niiden merkityksestä suhteessa toisiinsa sekä vaikutuksista onnistuneen visualisoinnin syntymiseen. Tarkoitus on selvittää, millaisten vaiheiden avulla tiettyyn tarkoitukseen kerätystä datasta pystytään jalostamaan ymmärrystä visuaalisuuden avulla. Mitä asioita tulisi ottaa huomioon, jotta datasta saisi muodostettua intuitiivisesti ymmärrettävissä olevia esityksiä, joilla merkityksiä ja informaatiota pystyisi välittämään valitulle kohderyhmälle? Jos kohderyhmään kuuluvat käyttäjät eivät pysty yksikäsitteisesti tulkitsemaan data ja siitä johdettua informaatiota, tulee visuaalisen tiedon hyödyntämisestä mahdotonta [Niemi, 2015].

Tutkimusmenetelmänä on perehtyä tiedon visualisoinnin ja sen lähialojen kirjallisuuteen ja tutkimuksiin. Tältä pohjalta pyritään luomaan synteesi tiedon visualisoinnissa tarvittavista erilaisista visualisoinnin prosessimalleista. Esitellyissä prosessimalleissa havaittuihin vahvuuksiin perustuen luodaan oma näkökulma visualisointiprosessiin. Kyseistä visualisointiprosessia havainnollistetaan ammattikorkeakoulun opintoasiainhallinnon yhteydessä soveltamalla sen eri vaiheita käytännön esimerkeissä. Tutkielmassa keskitytään abstrakteja asioita ja ilmiöitä kuvaavan datan visualisointiprosessiin.

Tutkielmassa visualisointiprosessin havainnollistamiseen käytetään ammattikorkeakoulun opiskelijoiden edellisen lukuvuoden opintasuorituskertymään liittyvää dataa. Siitä pyritään

nostamaan esiin merkityksellisiä ilmiöitä ja havaintoja visualisoinnin keinoin. Esimerkkivisualisointien pääkohderyhmäksi on ajateltu erityisesti opintoasiainhallinnon asiantuntijoita. Heidän kautta välillisesti tutkielman tuloksista voisivat hyötyä sekä ammattikorkeakoulun johto, tutkinto-ohjelmista vastaavat henkilöt, että muu ohjaushenkilöstö. Varsinaista lähdekoodia tai viimeisteltyä visualisointisovellusta tutkielmassa ei toteuteta, vaan painopiste on prosessin tarkastelussa kirjallisuutta sekä prototyyppejä apuna käyttäen.

Tutkielmassa syvennytään johdannon jälkeen tiedon visualisointiin liittyvään tutkimukseen sekä luodaan katsaus eri tieteenaloille, joissa on tehty visualisoinnin kannalta merkityksellistä tutkimusta. Luvussa kolme esitellään, vertaillaan ja analysoidaan erilaisia visualisoinnin prosessimalleja. Luvusta neljä lähtien visualisointiprosessin eri vaiheita käydään esimerkkiprosessin avulla läpi yksityiskohtaisemmin.

2. Tiedon visualisointi monitieteellisenä tutkimuskohteena

2.1. Visualisoinnin merkitys ja tutkimus

Tiedon visualisoinnin hyödyllisyys perustuu pohjimmiltaan ihmisen luontaisen näkö- ja havainnointikyvyn tehokkaaseen hyödyntämiseen tietojenkäsittelyllisiä menetelmiä käyttämällä. Ihminen pystyy käsittelemään visuaalisesti kuvien avulla samassa ajassa enemmän informaatiota kuin kaikilla muilla aisteillaan yhteensä. Esimerkiksi tekstejä lukemalla ihminen ei pysty sisäistämään niin suuria tietomääriä yhtä nopeasti kuin onnistuneen visualisoinnin avulla. [Ware, 2012]

Visualisoinnin avulla datasta pystytään intuitiivisesti havaitsemaan ilmiöitä, säännöllisyyksiä ja hahmoja, joita ei muussa muodossa esitetystä datasta välttämättä huomattaisi lainkaan [Ware, 2012]. Visualisoimalla ratkaistavia ongelmia niitä pystytään ymmärtämään paremmin ja voidaan tehdä parempia tietoon perustuvia ratkaisuja lyhyemmässä ajassa [Chittaro, 2006]. Datan määrän jatkuvasti lisääntyessä tiedon visualisoinnin merkitys ja tarve standardisoiduille, tutkimuksen toimiviksi todistetuille visualisointimenetelmille kasvaa. Tällaisten menetelmien avulla pystyttäisiin saamaan kerätystä datasta hyötyä toiminnalle riittävän tehokkaasti ja oikeanlaista tukea päätöksentekoon [Ware, 2012]. Visualisointi on väline tietomassoissa olevan sisällön hallintaan ja tulkintaan [Niemi, 2015].

Visualisoimalla on kuitenkin mahdollista myös johtaa käyttäjää harhaan ja levittää tahallisestikin väärää informaatiota. Visualisointien kehittämisprosessissa on mahdollista rajauksin tai näkökulmavalinnoin muuttaa olennaisesti visualisoinnin kohteena olevan ilmiön sanomaa, vaikka taustalla oleva data olisikin puolueetonta ja totuudenmukaista. [van Wijk, 2006 I] Klassikkokirjassaan “The Visual Display of Quantitative Information” Edward Tufte [1983] esittelee lukuisin esimerkein valhekertoimen (lie-factor) –käsitteen. Sillä hän pyrkii matemaattisesti kuvaamaan mahdollista vääristymää dataan perustuvan visualisoinnin ja varsinaisten datan arvojen välillä. Tämän takia onkin syytä muistaa, etteivät visualisoinnit sellaisenaan aukottomasti vahvista väittämiä totuudeksi, vaan myös visualisointien lukemisessa tarvitaan kriittistä tulkintataittoa.

Tiedon visualisointia ei edelleenkään kiistattomasti pidetä omana tieteenalana. Se on enemmänkin soveltavaa tiedettä, joka on syntynyt eri tieteenalojen tutkimustuloksia ei-formaalisti yhdistellen. Visualisoinnin asiantuntijaksi kehittyminen vaatiikin yhden yksittäisen tieteenalan teorioiden, viitekehysten ja menetelmien sisäistämisen sijaan syvällistä ja laajaa tietämystä useista erillisistä tutkimusalueista. Lisäksi tarvitaan kykyä huomioida kunkin yksittäisen visualisointitarpeen yksilölliset ja tapauskohtaiset erityispiirteet, sillä yhtä ainutta oikeaa tapaa toteuttaa visualisointi ei ole olemassa. Myös potentiaalisen käyttäjän kognitiiviset lähtökohdat määrittävät millainen

visualisointi olisi hänelle toimivin. Tästä syystä olisikin perusteltua mahdollistaa käyttäjille myös vaihtoehtoisia tapoja visualisointiin. Tällöin käyttäjä voisi itse valita niistä mieluisimman. [Niemi, 2015]

Tiedon visualisoinnin kannalta merkittävää tutkimusta on tehty mm. näkö tutkimuksessa (vision research), kognitiivisen psykologian alueella (cognitive psychology), HCI-tutkimuksessa (Human-Computer Interaction), tilastotieteessä (statistics), graafisen suunnittelun saralla (graphic design), kartografiassa (cartography) ja tietojenkäsittelytieteessä (computer science). Merkittäviä visualisointimenetelmien soveltajia löytyy näiden lisäksi esimerkiksi kauppatieteistä (business), fysiikasta (physics) ja lääketieteestä (medicine). Aloja, joilla visualisointimenetelmistä ei olisi tiedon hallinnan ja tulkinnan välineenä hyötyä, on huomattavasti vähemmän tai ei lainkaan käytettävissä olevan datan määrän jatkuvasti kasvaessa. [Hinterberger, 2009]

Visualisointeja voidaan hyödyntää tehokkaasti myös esimerkiksi datan laadun varmistamiseen. Oikeanlaisella visualisoinnilla esitettynä datassa olevat poikkeavat arvot erottuvat välittömästi ja havaitut virheet voidaan korjata joko datassa tai sen keräilytavassa. Yksi enenevässä määrin hyödynnetty käyttötarkoitus visualisoinneille on analyttinen. Lähtökohtana on muodostaa niiden pohjalta oletuksia tai päätelmiä, joita voidaan sen pohjalta ryhtyä tutkimuksellisin menetelmin tarkastelemaan ja todentamaan. [Ware, 2012]

Ammattikorkeakoulun toimintaympäristössä on monia erilaisia toimintoja, joissa tietoorientoituneita visualisointeja voitaisiin hyödyntää toimintojen tukemiseen tietoon perustuvalla tavalla. Toiminnasta syntyy jatkuvasti paljon dataa, jota systemaattisemmin hyödyntämällä voitaisiin synnyttää paljon arvokasta tietämystä jatkuvan kehitystyön ja päätöksenteon tueksi. Visualisointeja voitaisiin käyttää apuna myös datan puhtauden (oikeellisuuden ja puutteettomuuden) varmistamisessa sekä kohinan (esim. poikkeavien tai virheellisten arvojen) havaitsemisessa.

2.2. Näkö tutkimus ja visuaalinen havainnointi

Näkö tutkimus ja ihmisen havainnoinnin tutkimus psykologian alalla luovat visualisoinnille perustavanlaatuisen biologisen pohjan. Niiden tuotoksia voidaan pitää visualisoinnin kannalta kaikkein tärkeimpänä tutkimusalueena. Visualisoinneista saatavat hyödyt ovat käytettävissä ainoastaan näkö- ja havainnointikykyjä hyödyntämällä. Kognitiivisen psykologian alueella kehitettyjä menetelmiä sovelletaan erityisesti tiedon visualisoinnin evaluoinnissa. [Ware, 2012]

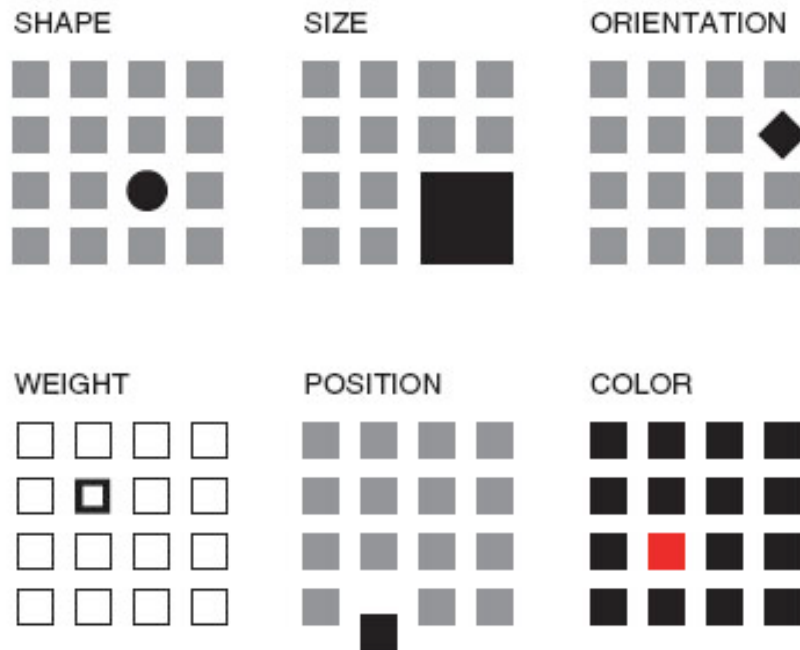
Tiedon visualisointi on näille tieteenaloille vuosi vuodelta merkittävämpi tutkimustulosten käytäntöön soveltaja. Visualisoinnin tarpeista syntyy myös uusia tutkimusaihoita, joihin paneutumalla pystyttäisiin entistä paremmin ymmärtämään esim. ihmisten kolmiulotteisuuden hahmotuskykyä tai muita havainnointiin liittyviä mekanismeja [Ware, 2012]. On tärkeää ymmärtää

sitä, miten ihmisen silmät, näköaisti, aivojen näköhavaintoihin keskittyneet alueet ja havainnointikyvyt toimivat. Näin pystytään luomaan mahdollisimman tehokkaita ja intuitiivisesti ymmärrettävissä olevia visualisointeja.

Ihminen havaitsee samanaikaisesti koko näkökenttensä alueella olevat valon ominaisuudet. Havainnointikykyään keskittämällä ihminen pystyy niiden pohjalta hahmottamaan luontaisesti niissä esiintyviä kuvioita. Havaituista kuvioista vain tehtävän tai kontekstin näkökulmasta relevantteina pidetyt objektit käsitellään näkömuistissa, joka pystyy käsittelemään yhdestä kolmeen objektia kerrallaan. [Ware, 2012] Ihmisen näköaisti työstää havaintoja siis käytännössä ensin alhaalta ylös, havaiten aluksi kaiken näkökentässä olevan, ja sitten ylhäältä alas keskittäessään huomionsa joihinkin tiettyihin elementteihin näkökentässään. Visualisoinnissa tämä tarkoittaa käytännössä sitä, että ensin graafisesta esityksestä havaitaan yleiskuva ja muodostuu jonkinlainen kokonaiskäsitely. Vasta tämän jälkeen huomio kiinnittyy tiettyihin erottuviin yksityiskohtiin esityksessä. Relevanteiksi koetut visuaaliset objektit täydentyvät mielikuvissa myös niiden ei-visuaalisilla piirteillä [Harvard, 2014], joten esimerkiksi tuttujen kuvien käyttämisen visualisoinneissa on osoitettu lisäävän niiden tunnistettavuutta ja muistettavuutta [Borkin et al., 2013].

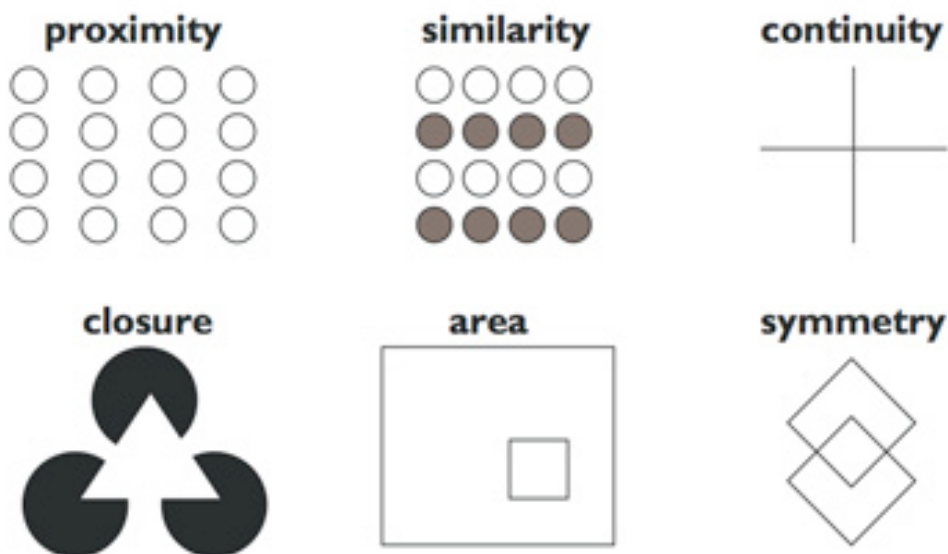
Näkötutkimus ja havainnointiin liittyvä tutkimus antavat konkreettisia suuntaviivoja käytännön visualisointityöhön. Esimerkiksi tulkintaa ohjaavat otsikot kannattaa aina sijoittaa lukussuunnan mukaisesti visualisoinnin alkuun, koska ihminen silmäilee visualisointeja samansuuntaisesti kuin lukee. Silmä hakeutuu selitteisiin ja muuhun tekstiin, joten ne kannattaa sijoittaa lähelle varsinaisia visuaalisia elementtejä. Näin ihmisen havainnointia ohjataan oikeaan suuntaan kohti visualisoinnin tärkeintä viestiä. [Harvard, 2014]

Toisena konkreettisena esimerkkinä voidaan mainita ihmisen luontainen kyky havaita kontrasteja. Riittävällä eroavaisuudella esimerkiksi visuaalisen objektin värissä, muodossa, tummuudessa, pituudessa, koossa, suunnassa, liikkuvuudessa tai ryhmittelyssä ihminen voi havaita jo alle 200-250 millisekunnissa joukosta poikkeavan elementin (kuva 3) [Ware, 2008]. Tätä ihmisiin sisäänrakennettua ominaisuutta hyödyntämällä voidaan nopeuttaa olennaisen erottamista epäolennaisesta ja ohjata käyttäjän huomio haluttuun visuaaliseen objektiin. Käänteisesti ajateltuna kontrastien (esim. värin, liikkeen) liiallinen käyttö puolestaan vaikeuttaa kokonaisuuden hahmottamista, jolloin visualisoinnin pelkistämää viestiä ei havaita.



Kuva 3. Esimerkkejä kontrastitekijöistä [Nanowerk, 2011].

Kolmantena esimerkkinä voidaan nostaa esiin ihmisen edistyneet hahmontunnistusmekanismit. Gestalt-psykologien 1920-luvulla lanseeraamat hahmolait ovat toimineet monella sovellusalueella reunaehtoina jo pitkään, mutta niistä voidaan hyötyä olennaisesti myös tiedon visualisoinnissa. Perusajatuksena Gestalt-teoriassa on se, että ihminen havaitsee ensisijaisesti osien summan pelkkien osien sijaan, mikäli hän hahmottaa ne intuitiivisesti yhteenkuuluviksi. Yhteenkuuluvuutta voidaan osoittaa visuaalisesti kuudella eri tavalla, jotka on nimetty Gestaltin laeiksi (kuva 4). Lait ovat läheisyys (proximity), samankaltaisuus (similarity), jatkuvuus (continuity), täydennettävyys (closure), alue (area) sekä symmetrisyys (symmetry).



Kuva 4. Hahmolait [firstcoastcreative, 2011].

Yleisimmin käytetyistä visualisointimenetelmistä esimerkiksi hajontakuviot (scatter plot) hyödyntää tehokkaasti ihmisen luontaista kykyä havainnoida arvopisteiden läheisyyttä ja jatkuvuutta, sekä sitä kautta myös niiden yhteenkuuluvuutta. Yleinen tapa esittää datan luokat eri väreillä luottaa ihmisen samanlaisuuden havaitsemiskykyyn ja esimerkiksi laatikkokuvaio (box plot) perustuu symmetriaan. [Ware, 2012]

Neljäs eri aloilla tehtyjen tutkimusten mahdollistama työkalu visualisointiin on värien käyttö. Visualisoinnin kannalta erityisesti värien havainnoinnin tutkimuksista on ollut paljon konkreettista hyötyä. Viestinvälityksen tehokkuuden kannalta värien harkitsemattomalla valinnalla voi saada aikaan paljonkin vahinkoa aikaiseksi. Toisaalta onnistuessaan sillä voidaan vahvistaa entisestään visualisoinnin välittämää viestiä. [Ware, 2012]

Ihminen havainnoi värien eroavaisuutta toisistaan paremmin, mitä suuremmalla alueella väriä on käytetty. Toisaalta pienillä alueilla käytettynä sävyerot jäävät helpommin havaitsematta. Tästä syystä visualisoinnissa elementtien on tarkoituksenmukaista olla riittävän suuria, mikäli niiden jokin ominaisuus on esitetty värillä. Lisäksi on pystytty osoittamaan, että ihminen havainnoi paremmin eroavaisuuksia tummuusasteiden muuttuessa musta-valkoskaalalla kuin eroavaisuuksia, jotka on esitetty pelkästään värisävyn muuttumisena. Tästäkin syystä värien käyttöä esimerkiksi datan arvojen vaihtelun kuvaajana kannattaa harkita tarkkaan. [Ware, 2012]

Datassa esitetyt arvot voidaan ryhmitellä niiden joidenkin ominaisuuksien perusteella luokiksi [Niemi, 2015]. Värien käyttöä suositellaan lähinnä tällaisten luokkien erottelussa. Luokkien erottelussa mustavalkoskaala ei yleensä riitä, sillä useimmin käytetyn valkoisen taustavärin ja mustan tekstin väliin jää ainoastaan kaksi toisistaan erottuvaa harmaata väriä, vaaleanharmaa ja tummanharmaa. Luokitteluvärejä käytettäessä parhaaseen lopputulokseen päästään käyttämällä vain toisistaan selkeästi erottuvia eri värejä, eikä esimerkiksi saman värin eri sävyjä. Suositelluimpina voidaan pitää maailmanlaajuisissa tutkimuksissa kaikkialla tunnetuiksi todettuja värejä eli mustaa, valkoista, punaista, sinistä, vihreää ja keltaista. [Ware, 2012]

2.3. HCI (Human Computer Interaction)

Tiedon visualisoinnin evaluoinnilla on vahvat juuret HCI:n tutkimusalueella, joka keskittyy tutkimaan ihmisen ja tietokoneen välistä vuorovaikutusta sekä kehittämään käyttäjäkokemusta. HCI:n tapaan myös tiedon visualisoinnissa pyritään ymmärtämään, miten visualisointi tukee tai voisi tukea ihmisiä heidän informaation tulkintaan liittyvissä tehtävissään. Kiinnostuksen kohteena on myös selvittää, miten ihmiset suorittavat tehtäviään, jotta pystyttäisiin toteuttamaan niiden tueksi entistä parempia järjestelmiä. [Carpendale, 2008]

Van Wijk [2006 I] vie HCI:n ja tiedon visualisoinnin yhteneväisyyden pisimmälle: hänen mielestään tiedon visualisointi on eräänlainen käyttöliittymän suunnittelutehtävä. Tästä johtuen

visualisoinnissa olisi tarkoituksenmukaista hyödyntää HCI:n suunnittelu- ja toteutusmenetelmiä. Näin esimerkiksi loppukäyttäjien huomioiminen sekä kehitystyön varhaisten prototyyppien käyttäminen saisivat suuremman painoarvon visualisointien suunnittelussa.

HCI:n käytänteitä ja teorioita hyödynnetään erityisesti tiedon visualisoinnin evaluoinnissa, sillä merkittävä osa erityisesti interaktiivisen tiedon visualisoinnin evaluoitavista piirteistä on samoja kuin HCI:ssakin. Tällöin keskiössä ovat vuorovaikutus käyttöliittymien kanssa ja käytettävyys [Carpendale, 2008].

Seuraavat järjestelmien evaluoinnille HCI:ssa määritellyt tavoitteet ovat lähes sellaisenaan sovellettavissa myös tiedon visualisointien evaluointiin:

- arvioidaan järjestelmän toiminnallisuus, eli tarkastellaan täyttääkö se kaikki sille asetetut toiminnallisuusvaatimukset;
- analysoidaan järjestelmän vaikutukset loppukäyttäjiin, eli tarkastellaan esimerkiksi graafisen käyttöliittymän käytettävyyttä, yksinkertaisuutta ja miten loppukäyttäjät sen hyväksyvät;
- tunnistetaan kaikki mahdolliset ongelmat, joita loppukäyttäjät voivat kohdata käyttäessään järjestelmää, kuten havainnoimalla toimintoja, jotka johtavat käyttäjiä harhaan tai tuottavat odottamattoman lopputuloksen. [Dix et al., 1998]

3. Tiedon visualisoinnin prosessi

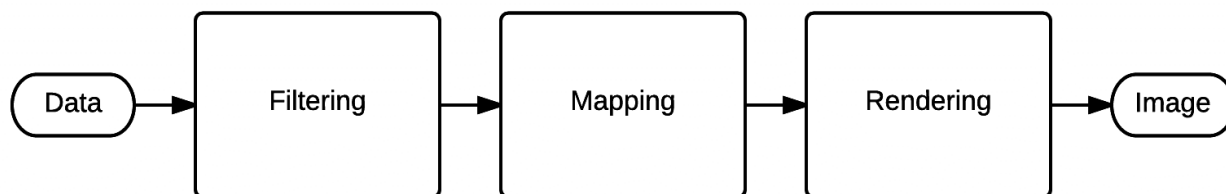
Tiedon visualisointi on vielä varsin nuori tutkimusalue ja mieluummin synteesi usean eri alan tutkimustuloksista ja menetelmistä kuin oma itsenäinen tieteenalansa. Sille ei ole vielä onnistuttu luomaan sellaista yhtenäistä teoreettista pohjaa ja prosessimalleja, jotka vanhemmilla ja vakiintuneimmilla tieteenaloilla toimivat kehityksen kivijalkana.

Visualisointia käsittelevässä kirjallisuudessa on keskitytty paljon erilaisten visualisointitekniikoiden kehittämiseen ja painopiste on usein visualisoinnin sovellutusorientoituneisuudessa [Niemi, 2015]. Muita usein toistuvia aiheita ovat olleet visualisointien luokittelu eri tavoin tai yksittäisen prosessin vaiheen käsittely muut vaiheet sivuuttaen. Pyrkimyksiä mallintaa tai kuvailla koko visualisointiprosessi kattavasti löytyy joitakin.

Seuraavissa alaluvuissa käydään läpi kirjallisuudessa esiteltyjä erilaisia visualisoinnin prosessimalleja ja tarkastellaan niiden näkökulmia ja rajoitteita.

3.1. Visualisointiputki

Ensimmäisiä ja yhä yleisimmin käytettävä visualisointiprosessin malli on Upsonin et al. [1989] ja Haberin ja McNabbin [1990] peräkkäisinä vuosina esittelemä visualisointiputki (visualization pipeline). Visualisointiputkessa visuaalinen esitys kehitetään datasta iteratiivisesti kolmen vaiheen lopputuloksena (kuva 5). Yhden iteraatiokierroksen lopputuloksen antamaa palautetta hyödyntäen voidaan aloittaa uusi visualisointiputken mukainen käsittely.

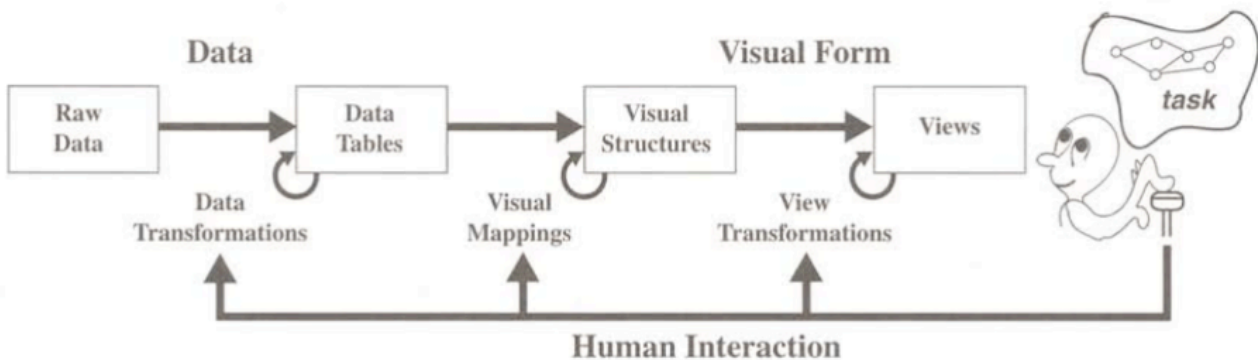


Kuva 5. Haberin & McNabbin [1990] visualisointiputki.

Suodatusvaiheessa (filtering) valitaan ja tarvittaessa rikastetaan käytettävissä olevasta datasta se datajoukko, joka on visualisoinnin kiinnostuksen kohteena ja jota visualisoinnin toteuttamisessa tarvitaan. Tämän jälkeen datasta valituille arvoille ja ominaisuuksille määritellään vastineeksi (mapping) jokin visuaalinen primitiivi, eli elementti (esim. piste, viiva) tai ominaisuus (esim. väri, sijainti, koko). Visuaalisten symbolien integroinnissa kokonaisesitykseksi (rendering) koostetaan valituista elementeistä ja ominaisuuksista visuaalisen esitys (image). [Haber & McNabb, 1990]

Alkuperäinen visualisointiputki keskittyy puhtaasti järjestelmässä tapahtuviin toimintoihin visualisointiprosessin kuvaamisessa. Se ei huomioi lainkaan käyttäjänäkökulmaa eli sitä, miten käyttäjä pystyy muodostamaan käsiteltävänä olevasta datasta ymmärrystä [Voigt et al., 2012]. Malli on kuitenkin merkittävä siksi, että se on ensimmäinen laajemmalle levinnyt visualisoinnin prosessimalli. Tästä syystä se on toiminut pohjana ja lähtökohtana useille muille prosessimalleille aina tähän päivään saakka.

Card et al. [1999] laajensivat visualisointiputkea huomioimaan niin käyttäjän tehtävän kuin vuorovaikutuksenkin merkityksen datan käsittelyn ja visualisoinnin luomisen eri vaiheissa (kuva 6).



Kuva 6. Visualisointiputken laajennos Card et al. [1999] mukaan.

Malli on onnistunut parannus alkuperäiseen visualisointiputkiesitykseen, koska näkökulmaa on laajennettu käyttäjän vaikutuksen huomioivaan suuntaan. Jalostetun mallin mukaan visualisointi on joukko iteratiivisia käyttäjän kontrolloimia muunnoksia, jotka pyrkivät tukemaan käyttäjää tehtävänratkaisussaan. Relevantti datajoukko esitetään mallin mukaan joukkona datatauluja (data tables). Varsinaiset datankäsittelyvaiheet ovat identtiset alkuperäiseen visualisointiputkeen nähden, joskin osa niistä on nimetty eri tavalla.

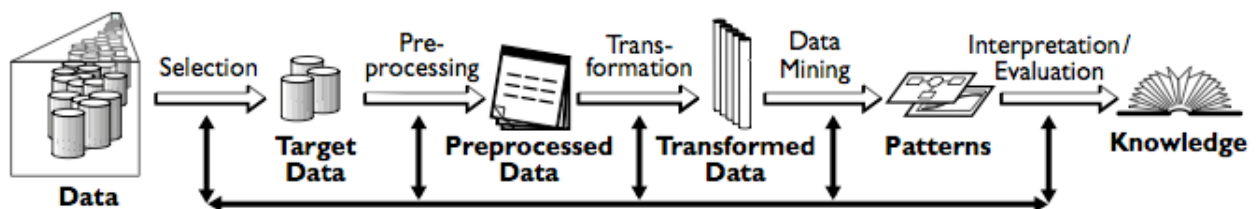
Esimerkkejä käyttäjän vuorovaikutuksesta visualisointiprosessissa on suodatusvaiheessa raja-arvojen ja muiden rajausten määrittely. Niiden pohjalta käyttäjä valitsee vain osan tietokannan tauluista tai tietueista mukaan visualisointiin. Datan ja visuaalisten primitiivien vastaavuuksien määrittelyvaiheessa käyttäjä voi vaikuttaa esimerkiksi siihen, käytetäänkö dataluokkia erottelevana tekijänä väriä vai muotoa. Visuaalisen esityksen koostamisvaiheessa käyttäjä voi yhä vaikuttaa lopputulokseen esimerkiksi muuttamalla horisontaalisen esityksen vertikaaliseksi. Näitä käyttäjän tekemiä visualisointikohtaisia valintoja voidaan kutsua visualisointiprosessin parametreiksi. [Jankun-Kelly et al., 2007]

Visualisointiputkea on kritisoitu visualisoinnin suunnittelijalle varsin armottomaksi menetelmäksi. Hänen on pystyttävä kontrolloimaan onnistuneesti jokaista datan muunnosvaihetta visuaaliseksi esitykseksi päästäkseen onnistuneeseen lopputulokseen. Tämä vaatii sekä syvällistä visualisoinnin

kohtealueen ja käsiteltävänä olevan datajoukon tuntemusta, että myös perehtyneisyyttä erilaisten visualisointimenetelmien mahdollisuuksiin ja rajoitteisiin. [Malyanov et al., 2013]

3.2. Tietämyksen muodostamisen prosessimalli

Tiedon visualisoinnin prosessissa on paljon samoja vaiheita kuin prosessissa, joka tähtää tietämyksen muodostamiseen tietokannoista (KDD, knowledge discovery in databases) (kuva 7). Lähtökohta ja tavoite ovat kummankin tutkimussuunnan prosesseissa samankaltaiset. Prosessi aloitetaan olemassa olevasta datasta ja sen päämääränä on käyttäjien tietämyksen lisääminen. Myös datalle tehtävät toimenpiteet lopputuloksen saavuttamiseksi ovat osin samanlaisia, kuten datan esikäsittely (preprocessing), muuntaminen (transformation), tulkinta (interpretation) ja evaluointi (evaluation). [Fayyad et al., 1996] Näiden vaiheiden lisäksi tietämyksen muodostamisen prosessissa on oleellista uuden informaation johtaminen (mining) analysoimalla saatavilla olevia tietoja [Niemi, 2015]. Tietämyksen muodostamisen prosessin osaamisesta on välitöntä hyötyä myös visualisointiprosessien ymmärtämisessä.



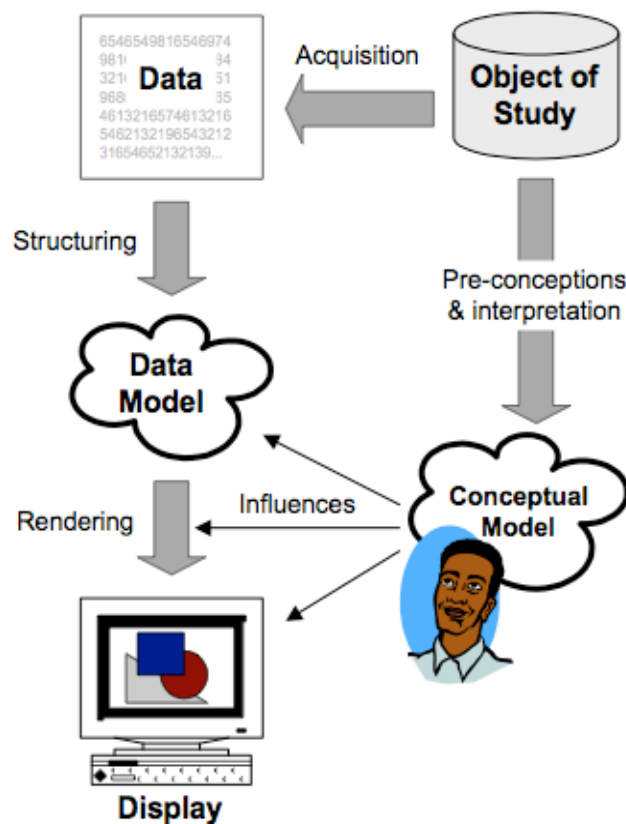
Kuva 7. Tietämyksen muodostamisen prosessi tietokantadatasta [Fayyad et al., 1996].

KDD tutkimusalueena keskittyy kokonaisvaltaisesti tietämyksen muodostamiseen ja löytämiseen datasta. Käytännössä KDD:ssä tutkitaan esimerkiksi, miten dataa säilytetään ja organisoidaan sekä miten siihen päästään käsiksi. Lisäksi KDD:ssä kiinnostuksen kohteena on tutkia, miten algoritmeja voidaan optimoida skaalautumaan isoihinkin datajoukkoihin tehokkuuttaan menettämättä sekä miten tuloksia voidaan tulkita ja visualisoida. KDD:ssä pyritään selvittämään myös, miten kokonaisvaltaista ihmisen ja tietokoneen välistä vuorovaikutusta voitaisiin mallintaa ja tukea. [Fayyad et al., 1996]

Tietämyksen muodostamisen tutkimuksessa visualisointi sisällytetään usein KDD-prosessin osaksi. Sen saama painoarvo on prosessissa kuitenkin vähäisempi kuin visualisoinnin tutkimuksessa, koska se ei ole ainut mahdollinen tapa esittää tietämyksen muodostamisen prosessin tuloksia käyttäjille. Toisaalta samalla tavalla tietämyksen muodostamisen prosessissa keskiössä oleva tiedonlouhintavaihe (data mining) voisi hyvin olla osa visualisointiprosessia. Visualisoinnin prosessissa se olisi yksi vaihtoehto rikastaa uutta merkityksellistä informaatiota raakadatasta visualisoinnissa käytettäväksi.

3.3. Toryn ja Möllerin malli

Melanie Toryn ja Torsten Möllerin [2002] visualisointimallin (kuva 8) lähtökohtana on tarkasteltavana olevaan ilmiöön tai asiaan (object of study) liittyvät tiedot. Tällainen ilmiö on Merriam-Websterin sanakirjan [2004] mukaan jokin fyysinen asia tai mielikuva, jota ajatellaan, jota kohtaan tunnetaan jotakin tai jonka suhteen toimitaan jotenkin. Koska kaikkia reaali maailman ilmiöitä ei ole tarkoituksenmukaista tai edes käytännössä mahdollista tutkia suoraan, tutkimuskohdetta tarkastellaan usein siitä kerätyn datan muodossa. Data määritellään siis tässä mallissa joukkona erilaisia arvoja, jotka on kerätty kuvaamaan jotakin ilmiötä tai asiaa, jota halutaan tarkastella välillisesti. [Tory & Möller, 2002]



Kuva 8. Toryn ja Möllerin [2002] näkemys suhteista visualisointiprosessin eri elementtien välillä.

Data esitetään ja organisoidaan jollakin tietomallilla (data model), johon perustuen siitä pystytään algoritmien avulla luomaan visuaalisia esityksiä [Tory & Möller, 2002]. Tietomallilla tarkoitetaan käytännössä sellaista systemaattista tietojen organisointitapaa, jolla esitetään datan varsinaisia arvoja, rakennetta, semantiikkaa, suhteita tai ominaisuuksia [Fayyad et al., 2002].

Visualisoinnin suunnittelijalla on aina jonkinlainen aiempaan tietämykseen ja tehtyihin esiselvityksiin perustuva käsitteellinen malli (conceptual model) kohdealueesta ja sen datasta. Hänen tulkintansa datan merkityksistä ja sisällöstä vaikuttavat olennaisesti sekä muodostettavaan

datamalliin, että sitä kautta myös visualisoinnin lopputulokseen. Hänellä on myös useita rooleja visualisointiprosessin aikana. [Tory & Möller, 2002] Tästä syystä onkin tärkeää, että visualisoinnin suunnittelija tekee koko suunnitteluprosessin ajan tiivistä yhteistyötä potentiaalisten käyttäjien kanssa. Yhteistyöllä voidaan saavuttaa visualisoinnin onnistumisen kannalta riittävä kohdealueen ymmärrys.

Toryn ja Möllerin [2002] malli ei rajoitu huomioimaan pelkästään suppeasti visualisointia tietojenkäsittelyn näkökulmasta lähtien datan valinnasta ja käsittelystä aina visuaaliseen esitykseen tai evaluointiin saakka, kuten monet muut mallit. Sen lähtökohdiana on järjestelmän ulkopuolella oleva tarkasteltava kohde. Mallissa huomioidaan vahvasti inhimillisten tekijöiden rooli ja vaikutus visualisointiprosessin kulkuun ja sen lopputulokseen. Vuorovaikutteiset visualisoinnit on jätetty tämän mallin ulkopuolelle.

3.4. Van Wijkin prosessimalli

Van Wijk [2006 I] esittelee visualisointiprosessiin käytännönläheisen lähestymistavan, joka on jalostettu visualisoinnin kontekstiin sovellettavaksi suoraan tuotekehityksen käytännöistä. Suunnittelutyön perussykli muodostuu kolmesta vaiheesta:

- tietosisältöön ja visualisointitapaan liittyvästä vaatimusmäärittelystä,
- ratkaisuvaihtoehtojen kehittämisestä sekä
- ratkaisuvaihtoehtojen arvioimisesta määrittelyjä vasten parhaan mahdollisen vaihtoehdon löytämiseksi.

Suunnitteluprosessi on iteratiivinen, joten suunnittelusykliä toistetaan useaan kertaan kohdealueen ymmärryksen lisääntyessä.

Vaatimusmäärittelyvaiheessa havaitut tarpeet tulisi pystyä kuvaamaan riittävän yksityiskohtaisesti ja jäsennellysti. Tavoitteena tulisi aina olla määrittelyt, jotka ovat sekä objektiivisia että mitattavissa olevia. Erilaiset tarpeet voivat vaihdella niin tärkeydeltään, painoarvoltaan kuin prioriteetiltakaan, joten myös nämä ominaisuudet tulisi kirjata tarvemäärittelyn yhteyteen. Epämääräiset määrittelyt vaikeuttavat ongelmanratkaisua sekä lopullisen parhaan ratkaisun valintaa prosessin myöhemmissä vaiheissa.

Van Wijk [2006 I] korostaa, että ensimmäisenä mieleen tuleva ratkaisu tai toteutusmalli ei välttämättä ole aina lopputuloksen kannalta paras. Tästä syystä ratkaisumalleja tulisikin kehittää useampia. Kaikkien ratkaisujen ei tarvitse olla uusia ja ennennäkemättömiä, vaan kaikenlaiset versiot sekä nykyisistä sekä uusista esitystavoista, että niiden yhdistelmistä ratkaisutapoina kannattaa ottaa vaihtoehtoina huomioon. On tärkeää myös selvittää, millaisia ratkaisuja käsiteltävään ongelmaan on jo kokeiltu tai olemassa, ja mitkä on koettu olevan niiden hyvät ja huonot puolet. Viimeisessä vaiheessa tarkastellaan eri ratkaisuvaihtoehtoja suhteessa alkuperäiseen vaatimusmäärittelyyn parhaan vaihtoehdon löytämiseksi. Tämä voidaan tehdä käytännössä

esimerkiksi ristiintaulukoimalla vaatimukset ja ratkaisumallit, sekä pisteyttämällä ne. [van Wijk, 2006 I]

Van Wijkin mallissa on otettu erinomaisesti huomioon visualisointiprosessin todellinen lähtökohta, eli kohdealueen ongelmien tunnistaminen ja jäsentely sekä tavoitteiden asettaminen prosessille. Lisäksi malli pyrkii välttämään järjestelmäkehitykselle tyypillistä huonoa käytäntöä, jossa hyväksytään ensimmäinen mieleen tuleva ratkaisutapa lopulliseksi toteutusmalliksi. Van Wijkin esittämä tuotekehitysmalli sopii mielestäni erinomaisesti myös visualisointiprosessin kehykseksi. Se ei kuitenkaan ota kantaa käytännön datan käsittelyyn, kuten visualisointiputkeen perustuvat prosessimallit.

3.5. Chittaron visualisointiprosessin tehtävälista

Luca Chittaron [2006] näkökulma visualisointiprosessiin on pikemminkin muistilista visualisoinnin vaiheista ja tehtävistä kuin varsinainen jäsenneily prosessimalli. Prosessin osien keskinäiset suhteet jäävät pitkälti lukijan oman tulkinnan varaan, huomioiden suunnitteluprosesseille tyypillinen iteratiivisuus. Tällainen korkeammalle abstraktiotasolle nostettu malli on toisaalta muita malleja joustavampi, koska prosessin vaiheita ei ole kiinnitetty tiettyyn suoritusjärjestykseen [Niemi, 2015].

Chittaron visualisointiprosessissa on kuusi tasoa, jotka hän esittelee tutkimuksessaan seuraavassa järjestyksessä. Vastaavuuksien määrittelyvaiheessa (mapping) päätetään millä visuaalisilla primitiiveillä kutakin datan piirrettä eli visualisointiin sisällytettävää kohdetta luonnehditaan. Valintavaiheessa (selection) ratkaistaan, mitkä piirteet datassa ovat relevantteja visualisointitehtävän kannalta. Esitystapavaiheessa (presentation) pohditaan, miten visualisointi tulisi toteuttaa, jotta se olisi käyttäjien laitteilla käsiteltävissä tarkoituksenmukaisella tavalla. Chittaro on tutkimuksessaan keskittynyt nimenomaan visualisointiin mobiililaitteille, joten esitystapavaihe on tästä johtuen hänen prosessissaan erityisen korostuneessa roolissa.

Vuorovaikutteisuusvaiheessa (interactivity) pohditaan, millaisia työkaluja visualisoinnin käyttäjälle toteutetaan. Miten käyttäjä pystyisi parhaalla mahdollisella tavalla hyötymään visualisoinnista ja muokkaamaan sitä? Prosessin lopputuloksen kannalta on tärkeää myös pohtia, onko visualisoinnissa huomioitu inhimilliset tekijät (human factors) eli ihmisen havainnointi- ja kognitiiviset kyvyt. Prosessin päätteeksi evaluointivaiheessa (evaluation) tarkastellaan visualisoinnin tehokkuutta loppukäyttäjien näkökulmasta.

Chittaron prosessin tasot on määritelty varsin käytännönläheisellä ja suppealla tasolla. Olettamuksia esitettyjen tasojen ulkopuolelta on useita, esimerkiksi visualisointiongelman on vain oletettu olevan olemassa ilman, että sitä on erikseen tutkittu tai määritelty prosessin alussa. Evaluointi keskittyy vain koko prosessin lopputuotokseen ja siinäkin pelkästään visualisoinnin tehokkuuteen. Evaluoinnissa jätetään huomiotta kokonaan esimerkiksi se, vastaako tuotettu visualisointi

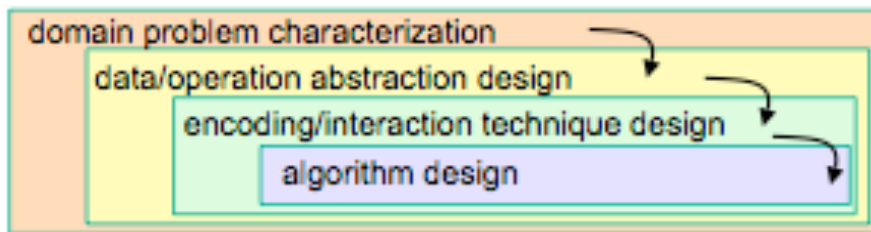
ongelmaan, johon se on tarkoitettu ja kuvaavatko valitut visuaaliset elementit parhaalla mahdollisella tavalla dataa. Inhimillisten tekijöiden vaiheen olisi mielestäni perustellusti voinut yhdistää osaksi evaluointia.

3.6. Munznerin sisäkkäinen prosessimalli

Tamara Munznerin [2009] visualisoinnin prosessimalli on selvästi Chittaron mallia perustellumpi ja jäsennellympi. Siinä on neljä sisäkkäistä tasoa (kuva 9):

- kohdealueen ongelman luonnehdinta (domain problem characterization),
- datan ja tehtävien abstraktin tason suunnittelu (data/operation abstraction design),
- dataelementtien visualisoinnin ja interaktiivisten toiminnallisuuksien suunnittelu (encoding/interaction technique design) sekä
- algoritmien suunnittelu (algorithm design).

Kaksi ensimmäistä vaihetta pyrkivät vastaamaan kysymykseen mitä visualisoidaan ja kaksi viimeistä kysymykseen, miten visualisointi toteutetaan [Niemi, 2015].



Kuva 9. Neljän sisäkkäisen tason malli [Munzner, 2009].

Ensimmäisellä tasolla visualisoinnin suunnittelijan on perehdyttävä aihealueen dataan ja käyttäjien tehtäviin niin, että hän pystyy ymmärtämään heidän tehtäväalueessaan käytettävää sanastoa ja työnkulkua. Toisella tasolla ensimmäisessä vaiheessa havainnoidut ongelmat ja kohdealueen data pyritään analysoimaan ja muuntamaan niin, että vaiheen lopputuloksena syntyy kuvaukset visualisointiin tarvittavista datatyypeistä sekä listaus tarvittavista toiminnoista. [Munzner, 2009]

Kaksi ensimmäistä tasoa ovat yleensä kaikkein vaikeimmat niiden abstraktin luonteen vuoksi ja käytännössä jäävät valitettavan usein vaille riittävää huomiota niin tutkimuksissa kuin käytännön visualisointiprojekteissakin. Tästä johtuen niin visualisointiprojekteissa kuin muissakin tietojärjestelmäprojekteissa tuotetaan valitettavan usein sovelluksia, jotka eivät todellisuudessa olekaan ratkaisuja olemassa oleviin ongelmiin tai vaihtoehtoisesti ne vastaavat sellaisiin ongelmiin, joita käyttäjillä ei ole. [Munzner, 2009] Tämä johtuu usein siitä, että käyttäjät eivät ole olleet riittävän tiiviisti mukana visualisointiprosessissa [Niemi, 2015].

Kolmas ja neljäs taso saavat yleensä eniten huomiota visualisoinnin tutkimuksissa ja toteutuksissa, sillä ne ovat vahvasti konkreettisia tasoja. Kolmannella tasolla suunnitellaan jo varsinaiset

visualisoinnin elementit ja vuorovaikutustavat ja neljännellä tasolla nämä toteutetaan algoritmitasolla. [Munzner, 2009] Valmista visualisointisovellusta käytettäessä algoritmien suunnitteluvaihe jää usein pois, koska algoritmeja on jo sisäänrakennettuna osaksi toteutustyökalua.

Munznerin [2009] visualisoinnin prosessimallin keskiössä on evaluointi. Mallin lähtökohtana on ajatus, että jokainen taso on syöte seuraavalle sisemmälle tasolle. Tästä johtuen kutakin tasoa on tärkeä evaluoida niin yksinään kuin yhdessä edeltävien tasojen kanssa, sillä jokaisella tasolla on omat erityispiirteet validoitavaksi. Tasojen välillä ei liikuta pelkästään lineaarisesti eteenpäin, vaan ymmärryksen lisääntyessä seuraavalla tasolla usein palataan iteratiivisesti kehittämään edellistä tasoa uuden ymmärryksen vaatimalla tavalla.

Edeltävällä tasolla tehdyt valinnat, ratkaisut ja myös virheet vaikuttavat merkittävästi seuraavien tasojen onnistumiseen ja toteutustapaan. Munzner [2009] toteaa varsin aiheellisesti, että erityisesti visualisoinnin tutkimuksissa on usein tapana käsitellä jotakin tai joitakin yksittäisiä tasoja, mutta sivuuttaa täysin sitä edeltävät tasot ja niistä aiheutuneet olettamukset. Tämä ilmiö on helposti havaittavissa useissa tässäkin tutkielmassa esitellyistä prosessimalleista. Munznerin mallissakin on tehty ainakin yksi vastaavanlainen olettaus: Käyttäjiä ei valita, tarkastella tai määritellä mitenkään. Prosessi aloitetaan suoraan kohdealueeseen, sen dataan ja tehtäviin tutustumisesta.

Vaikka Munzner [2009] itse kuvaa malliaan tahalliseksi yksinkertaistukseksi todellisuudesta, siinä on kuitenkin ansiokkaasti onnistuttu kuvaamaan useita tärkeitä näkökulmia visualisointiin. Näitä ovat esimerkiksi visualisoinnin eritasoisten vaatimusten hierarkia, keskinäiset vaikutussuhteet ja jatkuvan evaluoinnin tärkeys parhaan mahdollisen lopputuloksen saavuttamiseksi.

3.7. Yhtenäinen visualisointimalli

Yhtenäinen visualisointimalli (unified visualization model) on datan käsittelyyn ja vaiheittaiseen jalostumiseen perustuva malli (kuva 10). Malli koostuu viidestä tasosta ja neljästä muunnosvaiheesta näiden tasojen välillä. Alun perin mallin on luonut Martig et al. [2003], mutta sitä käsitellään tässä tutkielmassa Larrean et al. [2007] esittämästä näkökulmasta.

Mallin lähtökohtana on tarkasteltavan kohdealueen raakadata (raw data). Ensimmäisessä muunnoksessa se pyritään valintojen avulla supistamaan hallittavammaksi abstraktiksi dataksi (abstract data), jota ajatellaan tarvittavan visualisointiin. Tässä vaiheessa muodostetaan myös tarvittaessa metadataa raakadatan rinnalla tai sen sijaan visualisoinnissa käytettäväksi. Seuraavassa vaiheessa abstraktia dataa karsitaan tarvittaessa edelleen visualisoitavaksi dataksi (data to be visualized). Tämän jälkeen päätetään millä tavoin data visualisoitaisiin ja mikä visuaalinen elementti vastaisi parhaiten mitään visualisoitavan datan ominaisuutta (visual mapping). Viimeinen muunnos on visualisoinnin käytännön toteuttaminen (visualization transformation),

jonka tuloksena syntyy visualisoitu data (visualized data) käyttäjien tarkasteltavaksi. [Larrea et al., 2007]

Malli on erittäin datakeskeinen, eikä siinä oteta huomioon esimerkiksi datan keräilyä, datan kohdealuetta, ongelmamäärittelyä, inhimillisiä tekijöitä tai evaluointia. Mallissa käyttäjällä (user) tarkoitettaneen visualisoinnin suunnittelijaa eikä varsinaista potentiaalista loppukäyttäjää. Tämänkin mallin lähtökohtana voi päätellä olleen 90-luvun vaihteessa esitelty visualisointiputki, vaikkei sitä suoraan tutkimuksessa mainitakaan.

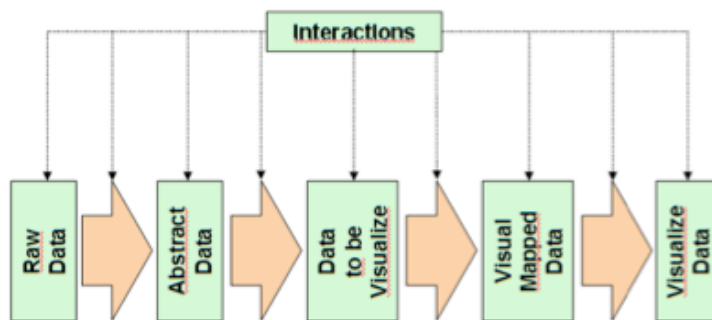


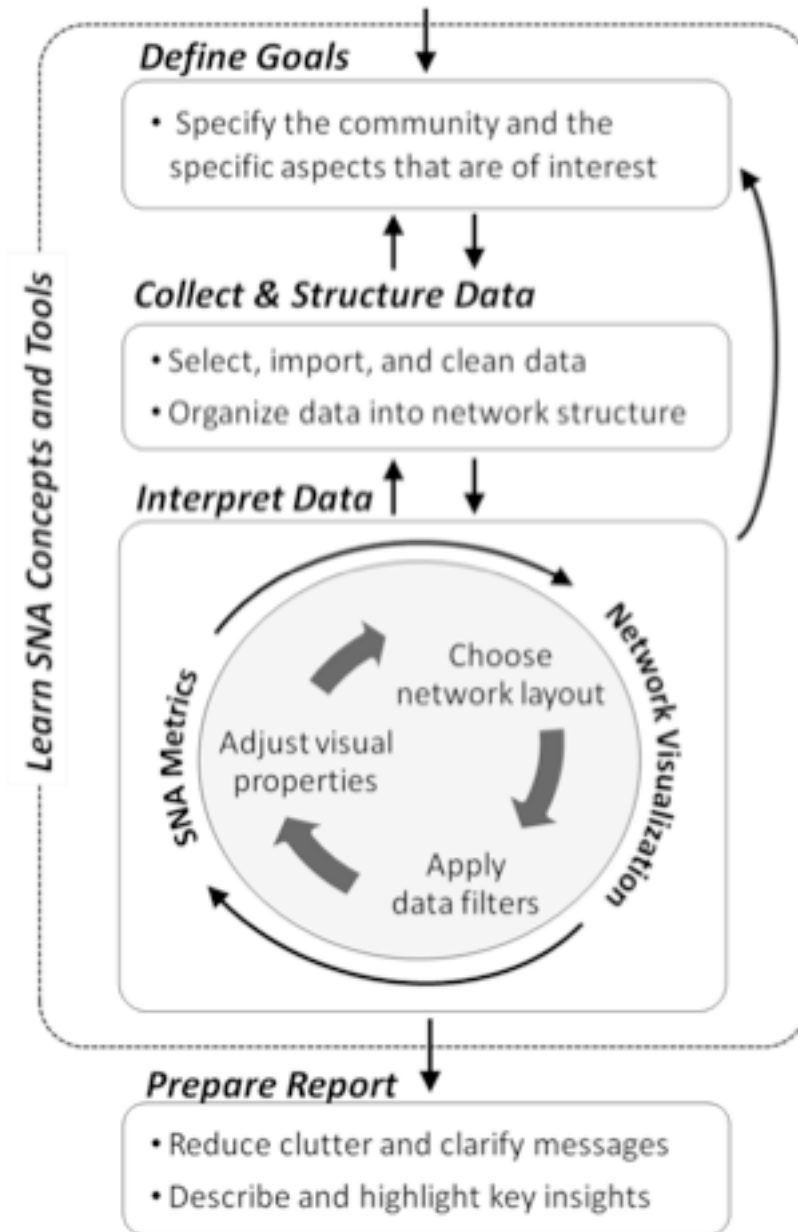
Figure 1. The visualization pipeline

The "Unified Visualization Model"	
Stages	Transformations
Stage "Raw data" Data from the application domain.	Transformation "Raw data to Abstract Data" This transformation allows the user to select the data he/she wants to visualize. After the selection, the data moves from the data domain representation to an inner and manageable structure.
Stage "Abstract Data" Data to be potentially visualize by the user. Besides this data the user also has the metadata created in the previous transformation.	Transformation "Abstract data to Data to be Visualize" From the "Abstract data" stage the user will select all the data that will be visualized.
Stage "Data to be Visualize" Data that will be visualized. It can be a subset of the "Abstract data"	Transformation "Visual Mapping" This transformation allows the user to specify how he/she wants to visualize all the data in the previous stage. All the necessary structures to support the spatial substrate, the visual elements and their attributes are created from this transformation.
Stage "Visual Mapped Data" Data to be visualized along with all the necessary information for its visual representation.	Transformation "Visualization Transformation" This transformation allows the creation on screen of all the data in the "Visual Mapped Data". This will usually include the application of some visualization technique that supports all the restrictions imposed in the "Visual Mapping" transformation.
Stage "Visualize data" This is the result from the visualization process. This is the starting point for the user to begin his/her visual exploration and navigation process.	

Kuva 10. Yhtenäinen visualisointimalli [Larrea et al., 2007].

Datan käsittelyn eri vaiheet on pilkottu mallin alkupäässä varsin pieniin kokonaisuuksiin, jolloin raja eri käsittelyvaiheiden sisältöjen välillä jää lähes keinotekoiseksi. Malli on aivan liian suppea ollakseen riittävä kuvaamaan koko visualisointiprosessia. Samalla se on myös liian tarkka, koska se herättää kysymyksen miksi prosessin ensimmäiset tasot ja muunnosvaiheet on eroteltu toisistaan. Ne voisivat yhtä hyvin olla yksi yhtenäinen taso ja yksi muunnos, jota toistetaan tarvittaessa iteratiivisesti.

3.8. Network Analysis and Visualization (NAV) –prosessimalli



Kuva 11. NAV-prosessimalli [Hansen et al., 2009].

Käytännönläheinen NAV-prosessimalli syntyi sosiaalisten verkostojen visualisointiin liittyvän tutkimuksen tarkastelusta. Prosessi perustuu koehenkilöinä toimineiden opiskelijoiden suorittaman

käytännön visualisointitehtävän havainnointiin ja siitä kerätyn aineiston analysointiin (kuva 11). Malli laajentaa tutkijoiden mukaan jo olemassa olevia sensemaking -malleja, joissa olennaista on iteratiivisuus datan keräämisessä ja analysoimisessa sekä asteittainen eteneminen kohti tietämystä lisääviä löydöksiä [Hansen et al., 2009]. Mallia pystyy yleistämällä soveltamaan myös muunlaisiin visualisointiprosesseihin, kun jättää verkostorakenteeseen viittaavat yksityiskohdat huomioimatta.

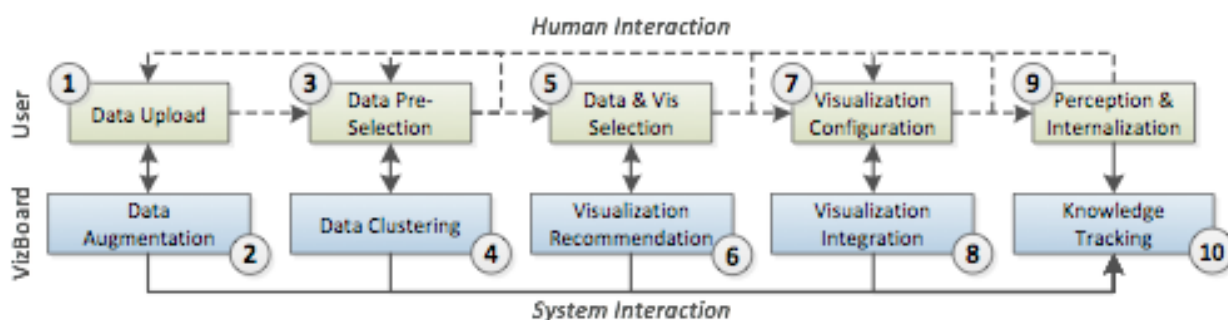
Prosessin lähtökohtana on määrittellä ensin prosessin tavoitteet (define goals). Toisin sanoen valitaan ja rajataan tutkittava aihealue sekä määritellään ne näkökulmat, joita ryhdytään tarkastelemaan. Tämän jälkeen kerätään valitun aiheen mukainen data (collect & structure data) ja esikäsitellään se tarkoituksenmukaisella tavalla. Esikäsiteltyä dataa tulkitaan (interpret data) käyttämällä valittua analyysimenetelmää (esim. laskemalla solmujen keskeisyyksiä tai keskeisyysasteita). Dataan yhdistetään sellaiset visuaaliset metaforat, joilla datan sisältämää merkitystä ja valittua näkökulmaa halutaan välittää visualisoinnin katselijalle. Lopputulosta vielä hienosäädetään (prepare report) poistamalla sieltä häiritseviä ja tarpeettomia elementtejä ja pyritään selkeyttämään visualisoinnin välittämää viestiä. Kaikkien vaiheiden rinnalle olennaiseksi on nostettu analysointitapaan liittyvien käsitteiden ja työkalujen opetteleminen (learn SNA concepts and tools) käytännön visualisointiprosessissa. [Hansen et al., 2009]

Hansenin et al. tutkimus on tehty varsin pienellä joukolla opiskelijoita, joille ei ollut määritelty etukäteen mitään tiettyä aihealuetta tai dataa, josta visualisointi tulisi tehdä [Hansen et al., 2009]. Tästä johtuen prosessin eteneminen on hieman erilainen kuin esimerkiksi todellisissa työelämän tilanteissa. Työelämässä aihepiirin valinta tai datan kerääminen eivät niinkään ole ne olennaisimmat työvaiheet, vaan aihepiiri on yleensä jollain tavalla ennalta rajattu. Myös visualisoitavaksi tarkoitettu data on usein jo valmiiksi kerätty johonkin muuhun käyttötarkoitukseen. Tavoitteenasettelukaan on harvoin suunnittelijan yksin määriteltävissä, vaan yleensä vaaditaan perehtymistä visualisoinnin käyttäjien ongelmiin ja toiveisiin. Myös lopputuloksen arviointi on jätetty kokonaan huomioimatta, mutta prosessin iteratiivisuuden malli tuo ansiokkaasti esiin.

Mallilla on kuitenkin oma arvonsa siitä näkökulmasta, että se ei perustu mihinkään ennalta määriteltyyn prosessimalliin, vaan syntyi käytännön opiskeluun liittyvän harjoituksen tekemistä tarkastelemalla. Malli validoi itsensä pienessä mittakaavassa jo syntyessään, joten oletettavaa olisi, että malli kestäisi myös isomman volyymin empiirisen testauksen. Lisäksi mallin ansiona voidaan muihin visualisointimalleihin verrattuna pitää sitä, että visualisoinnin työvälineen ja termistön käytön opetteleminen huomioidaan prosessin rinnalla kulkevana jatkuvana vaiheena. Nämäkin taidot väistämättä vaikuttavat visualisointiprosessin onnistumiseen, vaikkei niitä useimmissa prosessimalleissa vaikuttavina tekijöinä olekaan erikseen mainittu.

3.9. Visualisoinnin työkulkumalli

Voigt et al. [2012] ovat luoneet käyttäjän ja järjestelmän vuorovaikutuksen erottelevan työkulkumallin (kuva 12) erityisesti semanttisen verkon visualisointiin. Mallissa huomioidaan erityisesti informaatio sekä loppukäyttäjän tarpeet. Prosessi koostuu kymmenestä suurelta osin keskinäisessä vuorovaikutuksessa olevasta vaiheesta, jotka vuorottelevat käyttäjän ja järjestelmän välillä. Mallissa järjestelmänä on semanttisen verkon visualisointiin tarkoitettu VizBoard. Prosessissa Vizboardille määritellyt vaiheet voitaisiin suorittaa yhtä hyvin myös jollakin muulla visualisointityökalulla ja muuta kuin semanttisen verkon dataa käyttäen.



Kuva 12. Visualisointiprosessin työkulku Voigt et al. [2012] mukaan.

Visualisointiprosessi alkaa valitun raakadatan siirtämisellä järjestelmään käsiteltäväksi (data upload) [Voigt et al., 2012]. Mallissa ei huomioida visualisoinnin kohdealueen tai ongelman määrittelyä, eikä esimerkiksi datan keräilyä tai valintavaihetta lainkaan. Prosessi on päätetty aloittaa vasta pisteestä, jossa data on jo olemassa ja jollakin perusteella valittu visualisoitavaksi.

Järjestelmässä dataa rikastetaan (data augmentation) määrittelemällä onko jokin datan osa luonteeltaan nominaali-, ordinaali- vai määrällistä tyyppiä. Järjestelmä auttaa käyttäjää tässä toimenpiteessä näyttämällä erilaisia koostetietoja datasta ja tekemällä ehdotuksia käyttäjän hyväksyttäväksi tai muokattavaksi. [Voigt et al., 2012].

Datan esivalintavaiheessa (data pre-selection) käyttäjä saa nähtäväkseen karkealla tasolla datan rakenteen. Hän pystyy myös tutustumaan dataan tarkemmin vuorovaikutteisien menetelmien esimerkiksi zoomaamalla, etsimällä avainsanoja ja suodattamalla. Näillä menetelmillä käyttäjän olisi tarkoitus löytää kiinnostavia ilmiöitä visualisoitavaksi ja pystyä rajaamaan niiden visualisoinnissa tarvittava datan osajoukko. Esivalintavaiheen lisäksi järjestelmässä on algoritmeja datan semanttiseen klusterointiin [data clustering]. Tämä auttaa osaltaan käyttäjää löytämään datan relevantin osajoukon. [Voigt et al., 2012].

Seuraava vaihe on valita rajatusta datasta kiinnostavimmat muuttujat (data & vis selection), joiden perusteella järjestelmä ehdottaa visualisointiin parhaiten sopivia komponentteja käyttäjän

valittavaksi (visualization recommendation). Käyttäjän on mahdollista hienosäätää visualisointia (visualization configuration) esimerkiksi väritystä vaihtamalla tai vaikkapa vaihtamalla hajontakaavion asteikkoja. Konfiguroinnin jälkeen järjestelmä integroi (visualization integration) datan ja valitut visualisoinnit yhteen prosessissa määritellyllä tavalla käyttäjän havainnoitavaksi ja sisäistettäväksi (perception & internalization). Prosessin taustalla järjestelmä tallentaa tulevia visualisointiprosesseja varten tietoa käyttäjän tekemistä valinnoista ja ottaa vastaan myös suoraa palautetta käyttäjältä hänen tyytyväisyydestään prosessin tulokseen (knowledge tracking). [Voigt et al., 2012].

Esitetty malli on tässä tutkielmassa läpikäydyistä kaikkein käytännönläheisin ja yksityiskohtaisin omalla rajatulla prosessialueellaan. Se huomioi käyttäjän ja järjestelmän erillisinä toimijoina sekä heidän välisen vuorovaikutuksen yksittäisellä prosessin vaihetasolla. Mallista on varmasti hyötyä visualisointiprosessia suunnitteleville prosessin käytännönläheisimpiin vaiheisiin, joissa keskitytään nimenomaan visualisointijärjestelmän käyttämiseen, datan jalostamiseen ja visualisointimenetelmien valintaan.

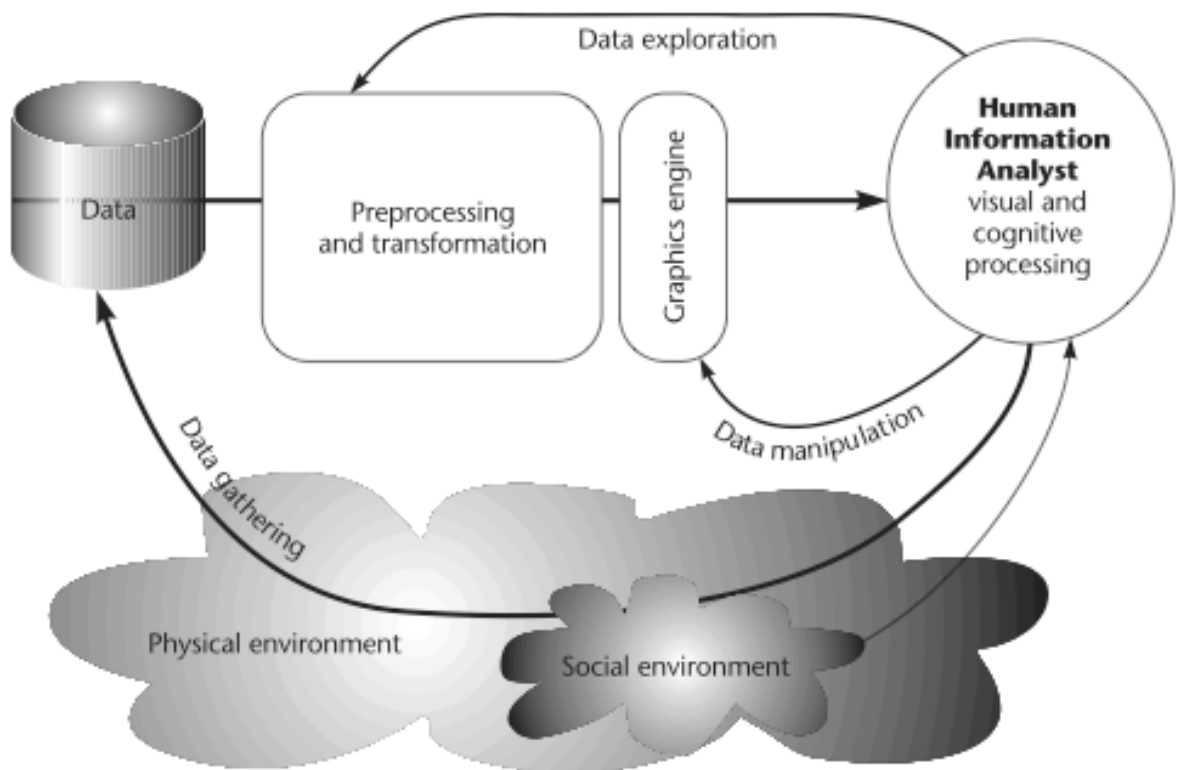
Koko visualisointiprosessin kuvaamiseen alusta loppuun saakka tätä mallia ei ole tarkoitettu, koska se alkaa vasta prosessin keskivaiheilta. Esimerkiksi prosessin alkuvaiheen määrittelyt ja tavoiteasettelut on jätetty pois, vaikka oletusarvoisesti ne tässäkin prosessissa tarvitaan, jotta olisi jotakin dataa järjestelmään siirrettäväksi ja jokin syy miksi visualisointia tehdään.

3.10. Waren prosessimalli

Colin Ware [2012] on esittänyt visualisoinnin prosessin nelitasoisena (kuva 13). Waren prosessin osia ovat:

- datan keräily ja tallentaminen,
- datan esikäsittely inhimillisesti tulkittavaan muotoon, joka
- näyttötekniikkaa ja graafisia algoritmeja hyödyntäen saatetaan visuaaliseksi esitykseksi näyttöruudulle, jota
- ihminen tulkitsee visuaalisten ja havainnointikykyjensä avulla.

Visuaalisen esityksen muotoon vaikuttavat niin keräilyvaiheessa tehdyt valinnat ja rajaukset kuin esikäsittelyvaiheessakin tehdyt linjaukset ja tulkinnat. Niitä täsmentämällä tai muuttamalla voidaan iteratiivisesti vaikuttaa visualisoinnin lopputulokseen ja siitä syntyviin johtopäätöksiin.

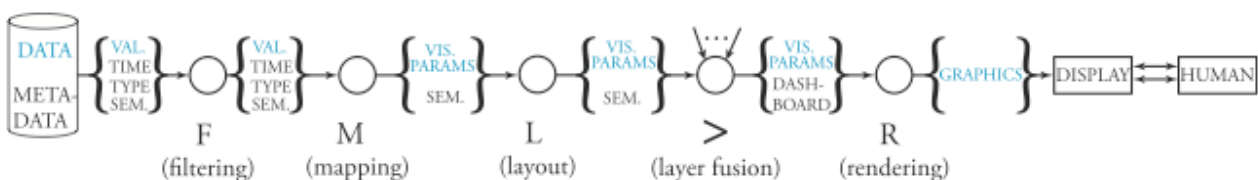


Kuva 13. Kaavio visualisointiprosessista Colin Ware [2012] mukaan.

Ware'n näkökulmasta prosessiin näkyy vahvasti hänen taustansa psykologian tieteenalalla. Ware'n malli huomioi kaikista tässä tutkielmassa läpikäydyistä malleista kaikkein mittavimmin ihmisen havainnointikyvyt sekä fyysiset ja sosiaaliset ympäristötekijät merkityksellisinä tekijöinä visualisointiprosessissa.

3.11. Malyanov et al. visualisointiprosessi

Malyanov et al. [2013] kuvaavat visualisoinnin prosessin jatkumona, joka alkaa raaka- ja metadatan olemassaolosta sekä päättyy visualisoinnin käyttäjän tietämyksen lisääntymiseen (kuva 14). Prosessi ei tietenkään ole näin suoraviivainen todellisuudessa, vaan jälkimmäisistä vaiheista saatu palaute palauttaa prosessin aiempiin vaiheisiin, joita kehitetään palautteen pohjalta iteratiivisesti.



Kuva 14. Visualisointiprosessi Malyanovin et al. [2013] mukaan.

Prosessimallissa on viisi datanmuunnosvaihetta (data transformation stages) sekä prosessin lopussa yksi vuorovaikutteinen vaihe (interaction stage). Tässä vuorovaikutteisuus tarkoittaa käytännössä

käyttäjäpalautetta loppukäyttäjältä kohdistettuina mihin tahansa prosessin edeltävään vaiheeseen [Malyanov et al., 2013].

Datanmuunnosvaiheista ensimmäinen on suodatus F (filtering), jossa data kerätään, suodatetaan ja lajitellaan. Data voi sisältää sekä tallennettua raakadataa, käyttäjiltä kerättyä dataa, että datasta johdettua metadataa tai tietämystä. Seuraava datanmuunnosvaihe on visuaalisten primitiivien valinta M (mapping) kuvaamaan kutakin visualisoitavaa datan piirrettä. Siinä valitaan esimerkiksi väri, sijainti tai muoto kuvaamaan tiettyä datan ominaisuutta. Asetteluvaiheessa L (layout) edellisessä vaiheessa muodostetut visuaaliset elementit sijoitellaan halutulle asteikolle tai alueelle. Yhdistämisvaiheessa > (layout fusion) asetellaan mahdolliset toisiaan tukevat visualisoinnit rinnakkain esimerkiksi tulostaulun (dashboard) muotoon. Viimeisessä vaiheessa tuotettu visuaalinen esitys koostetaan (R, rendering) lopullisessa esitysmuodossaan käyttäjien näytöille tarkasteltavaksi. [Malyanov et al., 2013]

Raakadatalle annettu semanttinen merkitys, joka käytännössä edellyttää kohdealueen asiantuntemusta, kulkee tärkeässä osassa mukana läpi koko prosessin. Prosessin alkuvaiheissa dataa käsitellään arvoina, mutta prosessin M-vaiheen jälkeen sitä käsitellään visuaalisina parametreina (visual parametri). [Malyanov et al., 2013] Käyttäjän antama palaute huomioidaan iteratiivisesti kaikissa eri vaiheissa. Tästä huolimatta esitetty prosessinäkökulma on varsin tekninen ja jättää huomiotta inhimilliset tekijät lopun interaktiivisuusvaihetta lukuun ottamatta. Nämä samat rajoitteet esiintyvät lähes kaikissa visualisointiputkimalliin perustuvissa variaatioissa ja laajennoksissa.

3.12. Yhteenveto ja pohdintaa visualisointiprosesseista

Visualisointiprosessia on eri tutkimuksissa lähestytty erilaisista näkökulmista ja prosessin eri vaiheita on nostettu keskiöön monenlaisin perusteluin. Vahvimpana vaikuttajana useissa malleissa näkyy suoraviivainen ja järjestelmäkokeskeinen visualisointiputki. Tietojenkäsittelyn alalla trendi on lisääntyvässä määrin pyrkiä käyttäjälähtöisyyteen ja ihmisen sisäistämisessä ja tulkinnassa tarvittavan kuormituksen minimoimiseen saatavilla olevan datan määrän lisääntyessä jatkuvasti. Tämä on hiljalleen alkanut näkyä myös visualisoinnin tutkimuksessa ja työkaluissa.

Tiedon visualisoinnilla pyritään helpottamaan käyttäjää hahmottamaan suuria datamääriä kerralla ja havaitsemaan niissä säännöllisyyksiä. Alkuperäiseen tavoitteeseen nähden visualisoinnin keinot ovat olleet pitkään todella järjestelmäkokeskeisiä, vaatien suurta visualisointimenetelmien ja kohdealueen asiantuntemusta. Prosessin vaiheissa ei ole pyritty käyttäjän toiminnan helpottamiseen ja kuormituksen minimoimiseen muualla kuin lopputuloksen tulkinnassa.

Syynä tähän saattaa olla se, että pitkään käytettävissä oli pelkästään staattisia visualisointeja, joita painettiin esimerkiksi lehtiin ja kirjoihin. Tällöin tehtävään perehtynyt asiantuntija kävi läpi visualisointiprosessin ja vain lopputuloksen täytyi olla laajemmille kohderyhmille helposti

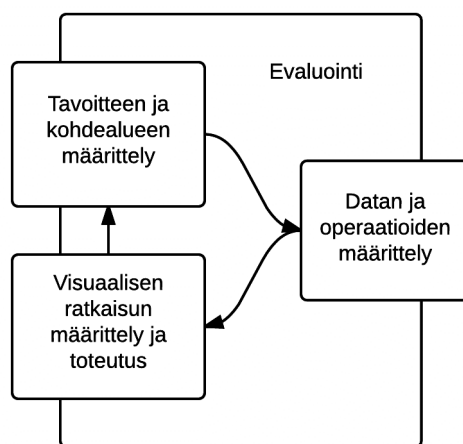
ymmärrettävässä muodossa. Vuorovaikutteisten visualisointien lisääntyessä tarvittaisiin prosessin kaikkiin vaiheisiin mahdollisimman paljon käyttäjäystävällisyyttä. Tällöin käyttäjät pystyisivät myös ilman visualisoinnin erityisasiantuntemusta tutustumaan oman asiantuntija-alueensa dataan ja muodostamaan siitä visuaalisia esityksiä myös muille.

Useiden visualisointiprosessien lähtökohtana on olemassa oleva data, jolloin datan valinnan perusteet ja visualisoinnin tavoitteenasettelu jäävät eksplisiittisen tarkastelun ulkopuolelle. Lähtökohtana oleva data on aina jollakin tavalla kerättävä tai rajattava sekä poimittava visualisointia varten alun perin muuhun tarkoitukseen kerätystä datasta. Jokaisella visualisoinnilla on myös jokin oletettu käyttäjäryhmä, päämäärä ja käyttötarkoitus, jotka vaikuttavat suoraan siihen, millaista informaatiota visualisoinnin välityksellä halutaan tuottaa. Näiden vaiheiden rajaaminen prosessimallien ulkopuolelle on varsin ongelmallista, koska niiden vähäisestä huomioimisesta voi aiheutua merkittäviä ongelmia visualisointien käyttökelpoisuuteen. Prosessin joka vaiheessa on oltava selvillä, mitä ollaan tekemässä, kenelle ja miksi. Myös evaluointi on ongelmallista, ellei ole täsmällisesti tiedossa, mihin visualisoinnilla on pyritty ja millainen sen tulisi olla.

4. Tutkielman lähestymistapa visualisointiprosessiin

Tässä tutkielmassa sovellettava visualisointiprosessi pohjautuu vahvasti em. Tamara Munznerin [2009] sisäkkäiseen prosessimalliin. Sitä sovelletaan erityisesti huomioiden HCI:n käyttäjakeskeistä näkökulmaa ja tuomalla mallin tueksi käytännön esimerkkejä mahdollisista käytettävissä olevista menetelmistä. Tutkielman mallilla pyritään luomaan käytännönläheinen lähestymistapa työelämälähtöisen visualisoinnin tarpeisiin ja sitä sovelletaan esimerkeillä ammattikorkeakoulun opintoasiainhallinnon yhteydessä.

Visualisointiprosessi koostuu kolmesta määrittelyvaiheesta sekä läpi koko prosessin tehtävästä evaluoinnista (kuva 15). Vaikka prosessin vaiheet tässä tutkielmassa on kuvattu lineaarisessa järjestyksessä, prosessin vaiheista palataan edeltäviin vaiheisiin iteratiivisesti aina tarvittaessa kohdealueen asiantuntemuksen ja visualisointispesifin osaamisen lisääntyessä. Visualisointiprosessi on kuitenkin syytä aloittaa aina tavoitteiden ja kohdealueen tietyn tasoisella määrittelyllä, koska kaikki seuraavat vaiheet pohjautuvat niihin. Evaluointi on elimellisesti mukana kaikissa prosessin vaiheissa, vaikka se esitelläänkin tässä tutkielmassa vasta muiden prosessin vaiheiden jälkeen.



Kuva 15. Näkökulma visualisointiprosessiin.

Tavoitteen ja kohdealueen määrittelyn vaiheessa luvussa viisi pyritään kuvailemaan millaisia ovat visualisoinnissa riittävän selkeät tavoitteet ja mihin tavoitteeseen tutkielman esimerkkiprosessilla pyritään. Kohdealuetta määriteltessä tunnistetaan ja valitaan kehitettävän visualisoinnin potentiaaliset käyttäjät, tutustutaan heidän tavoitteisiinsa, tehtäviinsä ja työskentelytapansa. Ensimmäisen vaiheen toteuttamisen menetelminä esitellään yhteistyömalleja.

Niiden avulla suunnittelija pystyy paremmin hahmottamaan oman roolinsa ja tavoitteensa suhteessa loppukäyttäjiin, jotta lähtökohta visualisointiprosessille olisi mahdollisimman realistinen.

Lisäksi käydään läpi varhaisten prototyyppien kehittämiseen perustuvaa menetelmää. Niiden avulla pystytään perehdyttämään kohdealueen asiantuntijoita visualisoinnin mahdollisuuksiin sekä syventämään visualisoinnin suunnittelijan osaamista kohdealueesta. Prototyyppien ja työpajojen tuella voidaan synnyttää uusia ideoita, joilla voitaisiin visuaalisesti tukea kohdealueen ongelmanratkaisua, kehittämistyötä ja päätöksentekoa.

Luvussa kuusi esitellään esimerkkivisualisointien taustalla oleva data, sekä keskitytään datan ja operaatioiden määrittelyyn. Siinä syvennytään erityisesti visualisoinnissa tarvittavan datan valintaan, analysoimiseen ja esikäsittelyyn. Tässä luvussa määritellään myös tarvittavat operaatiot, joiden avulla ensimmäisessä vaiheessa määritellyt tavoitteet voidaan kattavasti toteuttaa.

Visuaalisen ratkaisun määrittely- ja toteutusvaiheessa luvussa seitsemän valitaan visualisoinnin esitystapa sekä siinä esitettävälle luokille ja arvoille sopivat visuaaliset primitiivit. Visuaalisen ratkaisun määrittelyssä otetaan huomioon myös värien ja selitteiden käyttö sekä pohditaan operaatioiden toteutustapaa. Tässä luvussa esitellään myös esimerkkiprosessissa syntyneitä prototyyppisiä havainnekuvin.

Evaluointia tarkastellaan luvussa kahdeksan sekä yleisesti että kunkin yksittäisen vaiheen yhteydessä erikseen. Luvussa esitellään myös esimerkkiprosessissa toteutettu prototyyppien evaluointi. Evaluointi toteutettiin kohdealueen asiantuntijoiden avulla vapaamuotoisesti sekä yksittäisen käyttäjän havainnoinnilla ja ääneen ajattelun menetelmällä.

5. Tavoitteen ja kohdealueen määrittely

5.1. Tavoitteen määrittely

Hearst [2003] on määritellyt tiedon visualisoinnille seuraavat viisi yleistä tavoitetta:

- yhtenäistää isoja ja monimutkaisia tietojoukkoja,
- esittää tietoa useista näkökulmista,
- esittää tietoa useilta eri yksityiskohtaisuustasoilta,
- tukea visuaalista vertailua sekä
- mahdollistaa tarinoiden kerronta datasta.

Käytännön visualisointiprosessissa kuitenkin vaaditaan edellä esitetyn korkean tason yleisen tavoitteen lisäksi paljon yksityiskohtaista tietoa käyttäjiltä kohdealueen ratkaistavasta ongelmasta. Muun muassa tietoa tarvitaan datasta, nykykäytännöistä sekä visualisointikohteen käsitteistä, jotta visualisointiprojektille pystyttäisiin määrittelemään riittävän konkreettiset tavoitteet.

Van Wijkin [2006 I] mukaan ihanteellisesti määritellyt tavoitteet ovat

- kattavia, eli niistä ei jää puuttumaan mitään olennaista,
- erottelevia, eli erilaiset tarpeet on selkeästi eroteltu omiksi tavoitteiksi,
- objektiivisia, jolloin tavoitteiden toteutumista pystyy arvioimaan kuka tahansa sekä
- mitattavissa olevia, jolloin valitaan jo tavoitteenasettelussa ne mittarit, joiden avulla tavoitteiden toteutumista pystytään tarkastelemaan kiistattomasti.

Ideaalitavoitteita ei ole käytännössä helppo muodostaa varsinkaan visualisointiprosesseissa, koska tavoiteltu tietämyksen lisääntyminen on luonteeltaan varsin subjektiivista ja vaikeasti mitattavaa. Van Wijkin ihanteelliset tavoitteet ovat kuitenkin hyvä suunnannäyttäjä tavoitteita määriteltäessä, vaikka niiden täydelliseen kattavuuteen ei tähdättäisikään. [Van Wijk, 2006 I]

Koska visualisointia kohtaan asetetut vaatimukset voivat olla ja usein ovatkin keskenään ristiriidassa, on tärkeää liittää niihin kunkin tavoitteen tärkeyttä kuvaava painoarvo tai asettaa ne prioriteettijärjestykseen keskenään. Tavoitteiden riittävän yksityiskohtaisella määrittelyllä pystytään havaitsemaan visualisointiprosessin haasteet ja ongelmakohtat. [Van Wijk, 2006 I]

Tavoitteenasettelussa luodaan visualisoinnin alue, jonka sisällä seuraavissa vaiheissa ryhdytään kehittämään ratkaisuja [van Wijk, 2006 I]. Jos alue on määritelty väärin tai puutteellisesti, on vaikeampaa havaita, millaisiin ongelmiin visualisointia tarvitaan. Pahimmassa tapauksessa visualisoidaan ongelmia, joita ei ollut olemassakaan. Myös valmiin visualisoinnin evaluointi hankaloituu tai muuttuu käytännössä mahdottomaksi tietyiltä osin, ellei lopputulosta pystytä

vertaamaan siihen, mitä tavoiteltiin, tai arvioimaan onnistuiko visualisointi ratkaisemaan kohdealueeseen liittyvät alkuperäiset ongelmat.

Esimerkkitapauksemme on opiskelijoiden lukuvuosittaisen opintopistemääräkertymän visualisointi. Sillä pyritään välillisesti lisäämään 55 opintopistettä lukuvuodessa suorittaneiden opiskelijoiden määrää tulevaisuudessa, koska se on tällä hetkellä ammattikorkeakoulun toiminnan tehokkuuden yksi tärkeimmistä mittareista. Visualisoinnin tavoitteena on tuottaa useita erilaisia näkökulmia dataan, joiden pohjalta olisi helpompi havaita mahdollisia kehityskohtia. Niihin tarttumalla pystyttäisiin lisäämään 55 opintopistettä lukuvuodessa suorittaneiden opiskelijoiden lukumäärää. Menetelmiä kehityskohtien löytämiseen visualisoinnin keinoin voisivat olla esimerkiksi eri yksiköiden ja tutkinto-ohjelmien opintopistekertymäjakaumien keskinäinen vertailu sekä tarkastella käytetyn opiskeluajan suhdetta lukuvuoden aikana saavutettuun opintopistekertymään.

5.2. Kohdealueen määrittely

Visualisointi tehdään aina jostakin kohdealueesta, joka liittyy johonkin toimintaympäristöön. Kullakin toimintaympäristöllä on omat alalle tyypilliset käsitteensä, joilla kohdealueen ongelmia ja dataa kuvaillaan. Erilaisilla kohdealueilla on myös erilaiset käyttäjät sekä prosessit, joilla kohdealueen tehtäviä hoidetaan käytännössä ja joilla olemassa olevia ongelmia on tähän saakka pyritty ratkaisemaan. Onnistuneen visualisoinnin toteuttamiseksi suunnittelijan on tutustuttava sekä kohdealueen käsitteisiin, käyttäjiin että prosesseihin voidakseen ymmärtää sen dataa sekä käyttäjien tehtäviä ja tarpeita riittävällä tasolla. [Munzner, 2009]

Korkeakouluympäristön käsitteistä on koottu korkeakoulujen yhteistyönä yhteinen käsitteistö, jossa keskeisimmät käsitteet ja niiden keskinäiset suhteet on esitetty [Korkeakoulujen tietomalli]. Valmiiksi tehty käsittemäärittely helpottaa kohdealueeseen tutustumista merkittävästi ja nopeuttaa keskeisten käsitteiden merkityksen omaksumista. Kohdealuetta ohjaavilla tahoilla sekä korkeakoulun sisäisen toiminnan tukena on olemassa myös lukuisia muita ohjeita ja selvityksiä, joita tarvittaessa voi hyödyntää kohdealueeseen tutustumisessa. Kohdealueen määrittely koostuu käytännössä käsitteisiin ja tukimateriaaliin tutustumisen lisäksi käyttäjäanalyysistä ja tehtäväanalyysistä, joihin perehdymme tarkemmin seuraavissa alaluvuissa.

5.2.1. Käyttäjäanalyysi

On tärkeää mieltää jo prosessin alkuvaiheessa, minkälaisille käyttäjäryhmille visualisointia ollaan tekemässä. Täytyy ymmärtää mitä potentiaaliset käyttäjät haluavat ja mitä he visualisoinnin kohdealueesta tietävät. On myös tärkeää ottaa selvää, mitkä ovat käyttäjien pääasialliset tavoitteet visualisoinnin kannalta. Tarkoitus voi esimerkiksi olla säästää aikaa tai helpottaa oman työn tekemistä, lisätä asiantuntijuutta oman datan suhteen tai pelkästään säästää resursseja. [Hackos & Redish, 1998] Jos käyttäjien tarpeiden takana olevia vaikuttimia ei tiedetä, voi visualisointiprosessista tulla sellainen, joka toteutustavaltaan tai tuloksiltaan onkin vastoin

alkuperäistä tarkoitusta. Esimerkiksi useita vuorovaikutusmekanismeja sisältävä visualisointi, josta tulosten saaminen edellyttää käyttäjältä useita valintoja, ei palvele sellaista käyttäjää, joka tavoitteli oman työnsä helpottamista. Tällaisessa tilanteessa käyttäjät voivat kokea visualisoinnin käyttämisen niin hankalaksi, etteivät halua hyödyntää sitä omassa työssään.

Esimerkkimme visualisointiprosessissa käyttäjien määrittely ei ole helppoa eikä yksiselitteistä. Visualisointeja voitaisiin kehittää opintoasiainhallinnossa tuotettavien tilastojen ja taulukkomuotoisten tietojen rinnalle lisäarvoa tuottavaksi informaatioksi. Niitä voitaisiin tarjota myös johdolle, opetuksen vastuuhenkilöille ja muille hallinto- ja tukipalveluhenkilöstön edustajille. Opintoasiainhallinnon asiantuntijat ovat yksi keskeinen käyttäjäryhmä informaation välittäjän roolinsa lisäksi myös siksi, että he pystyvät omalla toimialueellaan reagoimaan joihinkin sellaisiin asioihin, joita visualisointien avulla voidaan saada esille. Opintoasiainhallinnon esityksestä voitaisiin esimerkiksi kehittää opiskeluoikeuteen tai erilaisiin kirjauskäytäntöihin liittyviä asioita.

Gonzálezin ja Kobsan [2003] tekemässä tutkimuksessa todettiin, että loppukäyttäjille on yleistä pitää visualisointeja ei-välttämättöminä lisätyökaluina muiden perinteisten menetelmien rinnalla eikä keskeisessä roolissa sellaisenaan. Tämä havainto näyttäisi pitävän jossain määrin paikkansa myös valitun kohdealueen työkuultuurissa ja potentiaalisten käyttäjien asenteissa koskien visualisointien käyttämistä. Visualisointeja ei ole kohdealueen toimintaympäristössä tähän mennessä käytetty kovin laaja-alaisesti. Niiden kehittämiseen tai visualisointityökalujen hyödyntämiseen ei ole myöskään erityisesti käytetty resursseja.

Erilaiset dataan perustuvat tilastot, yhteenvedot ja tiedoksiannot muille käyttäjäryhmille on perinteisesti tehty opintoasiainhallinnon asiantuntijoiden tekemin rajauksin ja määrittelyin heidän asiantuntemukseensa perustuen. Laajemman käyttäjäkunnan tarpeet ja toiveet on huomioitu vain, jos he oma-aloitteisesti antavat palautetta tai pyytävät jotakin tiettyä informaatiota. Systemaattisesti tarpeita ei ole toistaiseksi laajemmalla käyttäjäkunnalta kerätty. Koska opetus toteutetaan ja suunnitellaan aina muualla kuin opintoasiainhallinnossa, olisi perusteltua tehdä tulevaisuudessa syvällisempää tutkimusta myös visualisoinnin näkökulmasta erityisesti opetuksen parissa toimivien avainhenkilöiden tiedontarpeista.

Nämä lähtökohdat huomioiden on luontevinta myös visualisoinneissa käsitellä opintoasiainhallinnon asiantuntijoita potentiaalisina käyttäjinä, joilta kerätään tarpeita, rajauksia ja palautetta visualisointiprosessin eri vaiheissa. Jos näkemystä haluttaisiin kerätä laajemmalla käyttäjäjoukolta, täytyisi visualisointiprosessissa ja myös muissa dataan perustuvissa opintoasiainhallinnon tiedotus- ja tilastointiprosesseissa kehittää tulevaisuudessa uusi toimintamalli. Tällöin myös opetuksesta vastaavat henkilöt, johto ja muiden tukipalvelujen henkilöstö pystyttäisiin huomioimaan käyttäjinä jo suunnitteluprosessissa. Haasteeksi tässä muodostuisi mm. käyttäjäryhmän koko ja monimuotoisuus, erilaiset intressit ja yksilökohtaisen aktiivisuuden ja

sitoutuneisuuden erot. Tässä tutkielmassa tähän problematiikkaan ei syvällisemmin keskitytä, vaan käyttäjiksi määritellään opintoasiainhallinnon asiantuntijat.

Jos potentiaalisia käyttäjiä on vain vähän kuten esimerkkitapauksessamme, voidaan visualisointiprosessissa huomioida heidän yksilöllisiä luonteenpiirteitään tai arvojaan. Niillä voi olla merkitystä visualisoinnin hyödyntämisen kannalta. Jotkut käyttäjäryhmät voivat esimerkiksi arvostaa kekseliästä ja huumorilla rikastettua käyttöliittymää, kun taas toisille käyttäjille tärkeintä on perinteisempi virallinen ilme. [Hackos & Redish, 1998] Ammattikorkeakoulu on varsin perinteinen ja muodollinen toimintaympäristö ainakin henkilökunnan näkökulmasta, joten käyttäjät ovat tottuneet viralliseen ja koruttoman asialliseen esitystapaan. Tämä seikka on otettava huomioon myös visualisoinnin tyylissä.

Käyttäjien aiemmat kokemukset ja kertynyt osaaminen erilaisten työkalujen ja ylipäätään tietojärjestelmien käytöstä vaikuttavat siihen, minkälaisia visualisointeja heille kannattaa lähteä kehittämään. Vähäisillä tietoteknisillä perustaidoilla varustetulle kohderyhmälle painopiste suunnittelussa ja toteutuksessa on oltava helppokäyttöisyydessä, yksinkertaisuudessa ja selkeydessä. Käyttäjiltä vaadittavan opettelemisen ja muistamisen määrä on minimoitava. [Hackos & Redish, 1998]

Tällaisella kohderyhmälle suunnattuun visualisointiin kannattaa valita vain vähäinen määrä välttämättömiä toiminnallisuuksia ja attribuutteja. Visualisoinnin yhteydessä käytettyjen selitteiden, otsikoiden ja muiden tekstien sisältöä täytyy harkita huolellisesti, jotta visualisointi säilyisi riittävän yksinkertaisena ja intuitiivisena. Esimerkkikäyttäjäryhmässämme on eritasoisia tietojärjestelmäosaajia. Kaikilla on kuitenkin hyvät tietokoneen peruskäyttötaidot, joten visualisoinneissa voidaan hyödyntää myös vuorovaikutteisuutta tiettyssä määrin.

Visualisoinnin käyttäjäanalyysissa kannattaa tarkastella myös käyttäjien toimintaympäristöä, työskentelytapaa ja olosuhteita [Hackos & Redish, 1998]. Jos esimerkiksi käyttäjällä on taipumus käsitellä informaatiota lähtökohtaisesti paperilla eikä monitorin ruudulla, tai käyttäjäryhmällä ei ole sähköistä työpistettä käytettävissä, niin joukko staattisia visualisointeja on tällaisille käyttäjille oletettavasti parempi vaihtoehto kuin monipuolinen vuorovaikutteinen visualisointi. Myös painotuotteisiin, erilaisiin asiakirjoihin tai dokumentteihin tarvittavat visualisoinnit on järkevintä tuottaa staattisina kaavioina. Esimerkkitapauksemme käyttäjäkunnalla on tarvittavat sähköiset välineet käytössään. Mikäli visualisointeja olisi tarkoitus hyödyntää osana mustavalkoisia raportteja tai kokousmateriaaleja, tämä olisi huomioitava myös visualisoinnissa käyttämällä tummuudeltaan riittävästi toisistaan erottuvia värejä.

Perusteellisempaa datan tarkastelua tai analyysia tekeväälle asiantuntijalle rakennettu visualisointi voi olla ominaisuuksiltaan ja näkökulmavaihtoehtoiltaan moniulotteinen. Hänen tiedontarpeensa

on usein oletuksena muita käyttäjiä syvällisempi. Hän myös luultavasti käyttää visualisoinnin tarkasteluun enemmän aikaa kuin esimerkiksi joukko johtohenkilöitä, jotka vain vilkaisevat visuaalista esitystä kokouksessaan. Nämä kaikki erilaisiin olosuhteisiin ja työskentelytapaan liittyvät piirteet kannattaa huomioida mahdollisimman hyvin jo aikaisessa vaiheessa resurssien säästämiseksi.

Opintoasiainhallinnon asiantuntijoille tarjottavat visualisoinnit voivat sisältää joitakin rajausvaihtoehtoja ja erilaisia näkökulmia dataan. Niistä he voivat tarkastelunsa jälkeen muodostaa pienempiä kokonaisuuksia edelleen esitettäväksi tai vaihtoehtoisesti tarjota vuorovaikutteiset visualisoinnit sellaisenaan muillekin tarkemmasta informaatiosta kiinnostuneille käyttäjille.

Esimerkkivisualisoinneissa käytettävät luokittelevat tekijät ovat yleistettävissä myös muille käyttäjäryhmille. Tyypillisesti ammattikorkeakoulussa näkökulmat dataan ovat aina yksiköt, jotka ovat organisaatiohierarkiassa yleisin tarkastelutaso, ja tutkinto-ohjelmat, joiden välillä tehdään vertailua yleensä koosteissa ja tilastoissa. Opintopistekertymää pidetään yleisesti opiskelijan edistymisen mittarina kaikissa käyttäjäryhmissä.

5.2.2. Tehtäväanalyysi

Tehtäväanalyysin tarkoituksena on pystyä kuvailemaan kattavasti kohdealueen käyttäjien tehtäviä, niihin liittyviä alatehtäviä sekä menetelmiä, joilla näitä tehtäviä käytännössä suoritetaan. Tehtäväanalyysiä tarkastellaan tässä tutkielmassa kohdealueen määrittelyn näkökulmasta, eli sen tarkoituksena on tiedon kerääminen käyttäjistä ja käyttäjiltä heidän toimintansa ja tarpeidensa ymmärtämiseksi. Tehtäväanalyysissa suunnittelija pyrkii havainnoimaan tai muulla valitulla menetelmällä saamaan tietoonsa:

- käyttäjien tavoitteita ja
- miten he käytännössä pyrkivät näihin tavoitteisiin,
- miten he tehtäviään suorittavat ja
- miten reagoivat esimerkiksi ongelmatilanteissa. [Hackos & Redish, 1998]

Perusajatuksena visualisointiprosessin kannalta on löytää käyttäjien työkuluista vihjeitä siitä, millaisia ominaisuuksia ja toimintoja visualisoinnin tulisi sisältää ja minkälaisessa tehtäväkontekstissa visualisointia tulisi käyttää.

Tehtäväanalyysiä voi tehdä ainakin kahdella eri tasolla: tutustua käyttäjien ja järjestelmien toimesta tapahtuvaan työkulkuun yksittäisten prosessien näkökulmasta sekä perehtyä laajemmin käyttäjän toimenkuvaan. Työkulkunäkökulmassa tärkeää on havainnoida erityisesti kommunikointia prosessin eri vaiheissa ja toimijoiden välillä. Toimenkuvatarkastelussa tavoitteena on ymmärtää:

- miten usein käyttäjät tiettyjä tehtäviä suorittavat,
- miten tärkeitä tai vaikeita ne ovat,
- miten kauan niiden suorittamiseen kuluu aikaa sekä

- ovatko toimenkuvat kaikilla samassa tehtävässä toimivilla identtiset. [Hackos & Redish, 1998]

5.3. Menetelmiä

5.3.1. Yhteistyömallit

Visualisoinnin onnistumisen kannalta on äärimmäisen tärkeää, että visualisoinnin loppukäyttäjän ja suunnittelijan näkemykset ja tavoitteet kohtaisivat mahdollisimman hyvin jo prosessin alkuvaiheessa. Läheinen yhteistyö kohdealueen asiantuntijoiden kanssa on erittäin keskeistä, sillä kohdealueen ja visualisoinnin asiantuntijoilla on usein erilaiset osaamisalueet ja taustat. [van Wijk, 2006 II]. Visualisoinnilla tavoiteltu päämäärä on yhteinen, vaikka jaettua osaamista onkin vähän [Koh et al., 2011].

Kohdealueen datan käyttäjillä on aina ennakko-oletuksia käsiteltävästä datasta. Ne vaikuttavat olennaisesti heidän ymmärrykseensä siitä mikä on datan merkitys, mitä ilmiötä se kuvaa ja mikä olisi paras tapa visualisoida oletettu merkitys. Myös visualisointien suunnittelijoilla on olettamuksia käsiteltävästä datasta, ja visualisoinnit syntyvät heidän ymmärryksensä perusteella. [Tory & Möller, 2004]

Suunnittelijan on tärkeä pyrkiä kohdealueen asiantuntijoiden kannalta hyödylliseen lopputulokseen, eli sellaiseen visualisointiin, joka sopii heidän prosesseihinsa. Kohdealueen asiantuntijat käyttävät yleensä vain sellaisia työkaluja, joiden avulla he pystyvät tekemään työnsä entistä nopeammin tai paremmin. Van Wijk [2006 II] esittelee neljä erilaista käytännönläheistä mallia visualisoinnin suunnittelijan ja kohdealueen asiantuntijan yhteistyöhön. Useimmissa lähtökohtana on suunnittelijan samaistuminen visualisoinnin käyttäjiin. Mallit ovat ääriesimerkkejä, joita tarkoituksenmukaisesti yhdistelemällä päästään käytännössä parhaisiin tuloksiin.

Käyttäjäkeskeisessä suunnittelussa huomio kohdistuu loppukäyttäjien tarpeisiin, haluihin ja rajoitteisiin prosessin jokaisessa vaiheessa, ja käytettävyys on keskeisessä roolissa. Käyttäjäkeskeisen suunnittelun lähtökohtana on tiedettävä jo alkuvaiheessa, mitä käyttäjät ovat ja mitä he visualisoinnilta haluavat. [van Wijk, 2006 II] Suunnittelija tarvitsee varsin yksityiskohtaista tietoa kohdealueen ongelmista, prosesseista, työkuluista ja käsitteistä pystyäkseen tarjoamaan tehokkaita visuaalisia ratkaisuja niihin. Perehtymisvaihe vie paljon aikaa ja vaatii suunnittelijalta paljon paneutumista, jotta tarkoituksenmukainen visualisointi pystytään kehittämään. [Munzner, 2009]

Käytännössä aihealueeseen perehtyminen tällä tarkkuudella voi tarkoittaa esimerkiksi kohdealueen aihepiirin opiskelua, alan kirjallisuuden lukemista, loppukäyttäjien havainnointia työssään ja ideaalitulanteessa myös varsinaista osallistumista loppukäyttäjien jokapäiväiseen toimintaan.

Käyttäjakeskeisessä suunnittelussa prototyypit ovat tärkeitä. Niiden avulla visualisoinnin suunnittelija voi kehittää kohdealueen ongelmiin ratkaisuvaihtoehtoja ja käyttäjät arvioivat syntyneiden ratkaisujen käyttökelpoisuutta. Prototyypit auttavat saamaan palautetta suunnittelutyön tueksi, sillä suunnittelijan on käytännössä mahdotonta nähdä ja tulkita visualisointiaan samalla tavoin kuin loppukäyttäjän. [van Wijk, 2006 II]

Käyttäjakeskeisyys onnistuu vain, mikäli kohdealueelta löytyy asiantuntija tai useampia, jotka pystyvät sitoutumaan prosessiin ja antamaan siihen aikaansa. Prosessissa mukana olevien käyttäjien olisi hyvä olla avoimia uusille ideoille, näkökulmille ja toimintatavoille. Lisäksi valittujen asiantuntijoiden ongelmien tulisi kuvastaa laajemmin koko organisaation ongelmia tai kiinnostuksenkohteita, jotta visualisointia olisi kannattavaa lähteä toteuttamaan. [van Wijk, 2006 II] Aina tällaisia käyttäjiä ei kuitenkaan ole saatavilla [Koh et al., 2011].

Loppukäyttäjien näkökulmasta käyttäjakeskeisellä lähestymistavalla saadaan usein hyviä tuloksia. Visualisoinnin suunnittelijan intressejä kehittää jotakin uutta ja mielenkiintoista tulos ei välttämättä aina vastaa. Toisinaan loppukäyttäjän kannalta toimivin lopputulos voi olla visualisointina varsin tavanomainen, eikä vaadi erityistä innovatiivisuutta suunnittelijaltaan. Aina visualisointi ei tuota olennaista uutta tietämystä edes loppukäyttäjälle, vaikka visualisointiprosessi olisikin onnistunut. [van Wijk, 2006 II] Tällöin on tarpeen pohtia, olisiko jokin muu näkökulma hedelmällisempi vai onko todella niin, että datajoukko on esimerkiksi niin suppea ja hyvin ennalta tunnettu, ettei siitä enää löydy uutta kiinnostavaa tietämystä kohdealueen asiantuntijoille.

Käyttäjakeskeisyys vaatii paljon aikaa ja työtä sekä kohdealueen että visualisoinnin asiantuntijoilta [van Wijk, 2006 II], joten kovin edullinen ja nopea menetelmä se ei ole. Tästä johtuen käyttäjien ottaminen mukaan prosessin joka vaiheeseen ei kaikissa visualisointiprosesseissa ole resurssien puitteissa mahdollista tai edes kannattavaa [Koh et al., 2011].

Työkalusuunnittelussa visualisoinnin suunnittelija toimii käyttäjakeskeistä menetelmää passiivisemmin ja jättää omat ideansa ja luovuutensa taka-alalle. Tämän menetelmän lähtökohtana on loppukäyttäjä, joka esittää jonkin täsmällisen valmiiksi ideoidun tarpeen, jonka suunnittelija sitten sellaisenaan toteuttaa. Käyttäjän tarpeet voivat olla esimerkiksi jo olemassa olevien sovellusten ominaisuuksien yhdistelemistä, jolloin visualisoinnin suunnittelijaa ei käytännössä tarvita kuin teknisen toteuttajan rooliin. [van Wijk, 2006 II]

Tietojenkäsittelytieteen näkökulmasta visualisointiprosessiin voidaan ryhtyä myös kokonaan ilman loppukäyttäjien osallistamista. Tällä menetelmällä ei pyritä lisäämään loppukäyttäjien ymmärrystä, vaan esimerkiksi kehittämään jonkin olemassa olevan menetelmän ominaisuuksia tai resurssien tehokasta käyttöä. Lopputuloksista voivat hyötyä erilaiset käyttäjäryhmät välillisesti, vaikkei heitä suunnitteluprosessissa olekaan huomioitu. [van Wijk, 2006 II]

Suunnittelijakeskeisessä suunnittelussa suunnittelija valitsee kohdealueen ja käyttäjät oman mielenkiintonsa pohjalta. Tällöin suunnittelijalla usein on jo jonkinlaista aiempaa tietämystä kohdealueesta ja myös henkilökohtainen intressi pyrkii kehittämään ratkaisuja kohdealueen ongelmiin. Ihannetilanteessa myös muut tahot hyötyvät henkilökohtaisista kiinnostuksenkohteista lähteneestä visualisointityöstä, mutta yhtä lailla tulokset voivat jäädä merkityksettömiksi laajemmalle käyttäjäjoukolle, koska menetelmässä ei kerätä palautetta eikä sitä näin ollen myöskään saada. Yhdistettynä käyttäjälähtöiseen menetelmään tälläkin tavalla voidaan saada hyviä tuloksia, sillä suunnittelijan henkilökohtainen intressi on tehokas motivaatiota ja ongelmanratkaisun luovuutta lisäävä tekijä. [van Wijk, 2006 II]

Esimerkkitapauksessamme visualisoinnin suunnittelijalla on jo aiempaa asiantuntemusta kohdealueesta. Tämä helpottaa käytännössä yhteistyötä kohdealueen asiantuntijoiden kanssa, vähentää käyttäjakeskeisestä lähestymistavasta aiheutuvia kustannuksia, siihen tarvittavaa aikaa ja auttaa myös visualisoinnin tavoitteen määrittelyssä. Valittu yhteistyömalli esimerkkiprosessissamme on hybridimalli suunnittelijakeskeisestä ja käyttäjakeskeisestä mallista.

5.3.2. Tehtäväanalyysin menetelmiä

Tehtäväanalyysi pyrkii tuottamaan lopputuloksena listauksen visualisoinnin kohdealueen relevanteista tehtävistä sekä niiden edeltävyys-seuraavuus-suhteista. Analyysin tulos voi olla esimerkiksi:

- ajallisessa suoritusjärjestyksessä kirjoitettu lista tehtävistä,
- hierarkkinen esitys, jossa kukin tehtävä on pilkottu alitehtäviksi, tai esimerkiksi
- prosessikaavio [Hackos & Redish, 1998].

Tehtäväanalyysin toteuttamiseen voidaan soveltaa useita erilaisia laadullisen tutkimuksen menetelmiä joko yhdessä tai erikseen. Tyypillisimpiä ovat puolirakenteiset tai avoimet teemahaastattelut, kyselylomakkeet, havainnointi, käyttäjäpaneelit ja päiväkirjat [Ingwersen & Järvelin, 2005]. Tehtäväanalyysin tavoitteiden saavuttamiseksi esimerkiksi pelkkä käyttäjien haastattelu harvoin on kuitenkaan riittävän täsmällinen ja tarkka menettelytapa, vaan analyysiin tulisi hyödyntää myös muunlaisia menetelmiä [Munzner, 2009].

Tehtäväanalyysistä on hyötyä visualisoinnin tavoitteiden määrittelyssä ainoastaan, mikäli suunnittelija osaa poimia ja analysoida käyttäjien toiminnasta kerätystä datasta olennaisen informaation visualisointiprosessin tavoitteiden tueksi [Daabaj, 2000]. Tehtäväanalyysia onkin kritisoitu siitä, että se keskittyy liiaksi nykykäytäntöjen kuvaamiseen. Tämä johtaa helposti nykyisten järjestelmien ja toimintojen uudelleenrakentamiseen sen sijaan, että synnyttäisi uusia innovaatioita tai paljastaisi käyttäjien todellisia tarpeita [Richardson et al., 1996]. Tästä syystä tehtäväanalyysiä käsitellään tässä tutkimuksessa lähinnä kohdealueeseen tutustumisen ja

suunnittelijan ymmärryksen lisäämisen välineenä, ja vain välillisesti varsinaisessa tavoitteenasettelussa.

Tehtäväanalyysissa kannattaa ottaa huomioon myös se, että käyttäjät usein intuitiivisesti ilmaisevat haluavansa enemmän valinnanvaraa ja vaikuttamismahdollisuuksia kuin todellisuudessa tarvitsevat. Käytännössä vuorovaikutteiset toiminnot vaativat heiltä enemmän päätöksiä ja valintoja visualisointia käytettäessä. Tästä johtuen ne lisäävät merkittävästi käyttäjän kognitiivista kuormitusta. [Lam, 2008]. Tästä syystä tehtäväanalyysin avulla syntynyttä vaatimusmäärittelyä kannattaakin tarkastella kriittisesti sekä tässä että prosessin seuraavissakin vaiheissa. Toteutettavassa visualisoinnissa tulisi säilyä tasapaino helpon käytettävyyden ja monipuolisuuden välillä.

Esimerkkitapauksessamme opintoasiainhallinnon kohdealueella on jo valmiiksi olemassa paljon yksityiskohtaisia prosessikuvauksia kohdealueen eri prosesseista varsin yksityiskohtaisella tasolla. Myös asiantuntijoiden toimenkuvat on jäsennelty selkeästi. Niitä pystyttäisiin hyödyntämään laadullisten havainnointi- ja haastattelusessioiden tukena kohdealueeseen tutustumisessa, elleivät kohdealueen toimenkuvat, tehtävät ja prosessit ole entuudestaan tunnettuja.

5.3.3. Aikaiset prototyypit tavoitteenasettelussa

Yksi käyttäjäkeskeisyyteen perustuva toteutustapa tarpeiden, näkökulmien, ideoiden ja tavoitteiden keräämiseen käyttäjiltä on luoda jo prosessin aikaisessa vaiheessa prototyyppejä tai käyttää olemassa olevia sovelluksia erilaisten visualisointimahdollisuuksien esittämiseen [Koh et al., 2011]. Suhteellisen nopeasti luotavilla prototyypeillä sekä suunnittelijat että käyttäjät voivat tarkastella ja tarvittaessa täsmentää ja konkretisoida visualisointi-ideoitaan jo ennen kuin niiden perusteellisempaan luomiseen käytetään aikaa tai resursseja [Tory & Möller, 2004 II].

Visualisoinnin potentiaalisilla käyttäjillä ei välttämättä ole aikaisempaa osaamista visualisoinnista tai ymmärrystä siitä, mitä se mahdollistaa ja millaisia visuaalisesti tuettavissa olevia tarpeita heillä voisi olla. Vaikka staattiset visualisoinnit olisivatkin käyttäjille jo ennestään tuttuja, erityisesti vuorovaikutteisten visualisointien hyödyt, ominaisuudet ja mahdollisuudet ovat monelle epäselviä. Ne tulevat parhaiten esille prototyyppejä tai aiempia valmiita toteutuksia käytännössä kokeilemalla. [Koh et al., 2011]

Erilaisia visualisointeja voidaan esitellä loppukäyttäjille jo ennen kohdealueen määrittelyä ja siihen tutustumista. Esimerkkeinä voidaan käyttää joko olemassa olevia valmiita vuorovaikutteisia visualisointeja, joita löytyy esimerkiksi internetistä runsaasti, tai jostakin keinotekoisesta datasta nopeasti rakennettuja esimerkkivisualisointeja. Ensimmäisten esimerkkien ei tarvitse olla kohdealueen datasta tehtyjä, vaan ne voivat esittää minkä tahansa aihepiirin dataa vain havainnollistaakseen visualisointimenetelmää ja sen tarjoamia mahdollisuuksia. Tehokkaimmiksi

esimerkeiksi on kuitenkin koettu kohdealuetta lähellä olevien toimijoiden tarpeisiin tehdyt visualisoinnit. [Koh et al., 2011]

Aikaisten esimerkkien näyttämällä pyritään herättämään potentiaalisissa loppukäyttäjissä kiinnostusta, luomaan ymmärrystä visualisointien potentiaaliin ja käyttötapoihin sekä kannustamaan luovaan ajatteluun oman kohdealueen kontekstissa. Tavoitteena on, että käyttäjissä syntyisi konkreettisia tarpeita ja ideoita siitä, mitä heidän toimialueensa ilmiöitä voisi visuaalisin keinoin ymmärtää paremmin. [Koh et al., 2011]

Koh et al. [2011] toteuttivat aikaisten prototyypin esittämisen järjestämällä kohderyhmälle visualisointitietoisuustyöpajan (Visualization Awareness Workshop), jossa pyrittiin perehdyttämään potentiaalisia loppukäyttäjiä yleisiin visualisoinnin käsitteisiin. Työpajassa esiteltiin visualisointitekniikoita lähtien yksinkertaisimmista staattisista grafiikoista aina moniulotteisiin vuorovaikutteisiin visualisointeihin saakka. Työpaja koettiin erityisen hyvänä sellaisten käyttäjien kannalta, joilla ei ole aikaisempaa kokemusta visualisoinneista.

Suunnittelijan näkökulmasta visualisointitietoisuustyöpaja on erinomainen mahdollisuus kerätä uusia ideoita ja näkökulmia visualisointien käytännön merkityksistä loppukäyttäjiltä. Ilman prototyyppejä tai esimerkkejä visualisoinnin aihepiiriin ei pystyisi perehtymään yhtä tehokkaasti. Haasteeksi voi kuitenkin muodostua käyttäjien saaminen mukaan tällaisiin työpajoihin, koska niiden sovellettavuutta käytännön työhön tai merkitystä oman työn kehittämisen kannalta on vaikea etukäteen mieltää tai edes luvata. Työpajan osallistujilta, eli käyttäjiltä ja suunnittelijoilta, vaaditaan avoimuutta uusille ideoille ja menetelmille, luovuutta löytää uusia näkökulmia sekä aitoa kiinnostusta kohdealueen dataa kohtaan. Osallistujien motivaatiolla ja aktiivisuudella on iso merkitys siinä, muodostuuko työpaja hyödylliseksi työvälineeksi visualisointiprosessin tavoitteiden määrittelyvaiheessa, vai onko se vain kaikkien osapuolten resurssien tuhlaamista.

Kohdealueeseen tutustumisen ja sen ongelmien ja tarvemäärittelyn avuksi Koh et al. [2011] kehittivät toisenlaisen työpajan, jossa esitellään kyseisen kohdealueen datan visualisointiprototyyppejä (Domain Visualization Workshop). Työpajassa esitettävät visualisoinnit on luotu kohdealueen käyttämästä tai tuottamasta todellisesta datasta. Käyttäjät saavat kokeilla visualisointeja käytännössä ja antaa niistä palautetta suunnittelijoille. Visualisointien suunnittelijoilta nämä työpajat vaativat tietoisuustyöpajoja enemmän etukäteispanostusta. Useiden erilaisten visualisointien luominen kohdealueen datasta vaatii pohjaksi ymmärrystä kohdealueen käsitteistä, prosesseista ja datan semantiikasta sekä kykyä valita datasta potentiaalisesti kohderyhmää kiinnostavia näkökulmia visualisoitavaksi.

Prototyypin avulla havaitut mahdolliset virhetulkinnat tulevat käyttäjien palautteesta työpajassa välittömästi ilmi ja suunnittelijoiden kohdealueen ymmärrys paranee. Onnistuneet prototyypit

voivat parhaimmillaan toimia myös lopullisten toteutusten pohjana, jolloin tähän vaiheeseen käytetty työpanos voidaan säästää tulevilla työvaiheilla. Työpajaan valmistautuminen on suunnittelijoille käytännönläheinen tapa perehtyä kohdealueeseen ja sen dataan sekä saada potentiaalisilta loppukäyttäjiltä palautetta ja uusia ideoita. Loppukäyttäjille syntyy parhaimmillaan mahdollisuus tehdä aikaisia löydöksiä omasta datastaan. Vähimmilläänkin loppukäyttäjien tietämys visualisoinnin mahdollisuuksista omalla toimialueella lisääntyy ja näkemys konkreettisista visualisointitarpeista voi jäsentyä. [Koh et al., 2011]

Näiden kohdealueen jäsentämiseen ja visualisointitarpeiden löytämiseen pyrkivien työpajojen onnistuminen on kiinni useista eri tekijöistä. Suunnittelijoilla on merkittävä rooli esiteltävien prototyyppien luomisessa. Jos datasta on onnistuttu valitsemaan oikeanlaisia ominaisuuksia ja ulottuvuuksia prototyyppiin, käyttäjät mieltävät prototyypit helpommin hyödyllisiksi. Tällöin prototyypit toimivat täsmällisempien tarvemäärittelyjen ja uusien ideoiden pohjana. Jos prototyypit eivät kerro käyttäjille kohdealueesta mitään merkityksellistä tai jäävät liian kauas käyttäjien arkitodellisuudesta, visualisoinnit voidaan yleisestikin kokea merkityksettömiksi omalla toimialueella. Myös käyttäjien aktiivisuudella, kiinnostuksella, avoimuudella ja innovatiivisuudella on suuri merkitys työpajan lopputulokseen. Työpajojen tulokset ovat luonteeltaan laadullisia, eikä niitä suoraan voi hyödyntää muissa visualisointiprosesseissa tai yleistää koskemaan laajempia käyttäjäryhmiä tai toimialueita.

Uusien visualisointi-ideoiden ja -tarpeiden synnyttämiseksi ammattikorkeakoulun toimintaympäristössä voitaisiin laajemminkin soveltaa näiden kahden työpajan sisältöjä vapaamuotoisesti yhdistellen. Tällaisissa työpajoissa esimerkiksi muutamaa ennalta valittua kohdealueen asiantuntijaa voitaisiin käyttää apuna visualisointi-ideoiden kehittämisessä ja prototyyppien arvioimisessa jatkokehitystä varten. Kohdealueen datasta voitaisiin luoda käyttäjien kokeiltavaksi valmisohjelmistoa käyttäen useita erilaisia visualisointeja. Tällöin dataan saataisiin luotua useita erilaisia näkökulmia, ja visualisointien ideoita sekä mahdollisuuksia pystyttäisiin välittämään käyttäjille jo aikaisessa vaiheessa.

Tutkielman esimerkin yhteydessä syntyneitä prototyyppejä potentiaalisille käyttäjille esittelemällä ja evaluoimalla onnistuttiinkin synnyttämään useita uusia ideoita liittyen kohdealueen muihin toimintoihin. Niissä visualisoinneista voisi olla tulevaisuudessa hyötyä prosessien ja käytäntöjen kehittämisen apuna. Ilman konkreettisia prototyyppejä visualisoinnin potentiaalisten hyötyjen ja mahdollisuuksien välittäminen onnistuneesti kohdealueen asiantuntijoille olisi ollut haastavampaa.

Koska tähän tutkielmaan ei varsinaista käyttöön otettavaa visualisointitoteutusta sisällytetä, käyttäjiä edustavat asiantuntijat saivat nähtäväksi ja arvioitavaksi valmisohjelmistolla tehtyjä aikaisia prototyyppejä suunnitelluista visualisoinneista. Palautteen perusteella niitä voidaan iteratiivisesti kehittää tarkoituksenmukaisemmiksi ennen kuin niitä ryhdytään toteuttamaan varsinaisessa

toimintaympäristössä käytettävissä olevilla työkaluilla. Tällä menetelmällä varmistutaan, että visualisoinnit todella koetaan apuvälineiksi kohdealueen ongelmiin, ennen kuin niiden rakentamiseen käytetään resursseja suuremmissa määrin.

6. Datan ja operaatioiden määrittely

Visualisointiprosessin tässä vaiheessa kohdealueen arkikielellä kuvaillut ongelmat ja käsitykset datasta pyritään muuttamaan tietojenkäsittelyn näkökulmasta ymmärrettäviksi ongelmiksi ja käsitteiksi. Tällöin niitä voidaan käyttää pohjana visualisointityön seuraavissa vaiheissa. [Munzner, 2009] Dataa täytyy analysoida edellisessä vaiheessa kerätyn ymmärryksen pohjalta siten, että siitä pystytään valitsemaan tavoitteiden kannalta olennaiset attribuutit. Lisäksi on määriteltävä näihin attribuutteihin liittyvät datatyypit, jotta seuraavassa vaiheessa pystyttäisiin valitsemaan kullekin attribuutille sopivin visuaalinen primitiivi.

Data muodostuu tarkastelukohteena olevista tapauksista tai tapauksia kuvaavien ominaisuuksien arvoista [Tan et al., 2006]. Esimerkiksi opiskelijoiden lukuvuosittaista edistymistä kuvaavassa datassa tarkasteltava yksittäinen kohde on opiskelija ja opiskelijasta kuvattuja ominaisuuksia ovat mm. käytettyjen läsnäolokausien määrä, erilaiset opintopistekertymät eri ajanjaksoilla sekä opiskelijan tutkinto-ohjelma. Näille ominaisuuksille annetut arvot vaihtelevat opiskelijakohtaisesti.

Opintoasiainhallinnon kohdealueella on olemassa paljon muutakin dataa, joka voisi tukea 55 opintopisteen lukuvuodessa suorittaneiden tarkempaa tarkastelua. Tässä tutkielmassa siitä on käytettävissä vain ennalta rajattu koosteinen otos. Tästä syystä tässä visualisointiprosessin sovelluksessa ei päästä soveltamaan laajasti esimerkiksi datalähteiden valintaa tai datan yhdistelemistä useista eri lähteistä. Samasta syystä prosessin ensimmäisessä vaiheessa määriteltyä tutkimusongelmaa lähestytään visualisointiprosessin esimerkkisovelluksessa varsin datalähtöisesti. Käytettävissä olevasta datasta etsitään ominaisuuksia, joita visualisoimalla pystyttäisiin parhaiten tukemaan alussa määriteltyä tavoitetta.

Datan ja tehtävien määrittely sisältää käytännössä visualisointiputkimallisissa kuvatun suodatusvaiheen (filtering). Olennaiset dimensiot datasta valitsemalla pystytään visualisoimaan opiskelijoiden edistymistä ensimmäisen vaiheen edellyttämällä tavalla. Vaiheen tarkoitus datan käsittelyn näkökulmasta on muuttaa ja karsia raakadatan joukkoa suppeammaksi poimimalla siitä edellisen vaiheen tuloksena saadut olennaiset ominaisuudet. Lisäksi datan ulottuvuuksia voidaan myös vähentää tai lisätä esimerkiksi johtamalla datasta uusia arvoja metadataksi [Upson et al., 1989]. Dataa esikäsitellään, jotta se soveltuisi paremmin analysoitavaksi ja jatkokäsiteltäväksi visuaalisin keinoin.

Tässä tutkielmassa käytetty esimerkkidata sisältää joitakin valmiiksi johdettuja arvoja, kuten opiskelijan opintopistemääräkertymän edellisenä lukuvuonna ja edellisen lukuvuoden lukukausina sekä käytettyjen läsnä- ja poissaolokausien lukumäärän. Data tässä muodossa on peräisin

korkeakoulujen valtakunnallisesta tietovarannosta eli VIRTIA-opintotietopalvelusta [Husa, 2015]. Se on muodostettu VIRTIA-opintotietopalveluun ladatusta opintosuoritusten, opiskeluoikeuksien ja opiskelijoiden datasta tietyin laskentamenetelmin. Käytännössä data on siis korkeakoulun perusrekisteridataa, joka on ensin ladattu VIRTIA-opintotietopalveluun ja josta on johdettu tiettyjä summa-arvoja.

Opiskelijoihin liittyvä data on kerätty alun perin muuhun tarkoitukseen, joten sitä ei sellaisenaan ole optimoitu visualisointitarkoituksiin. Esimerkkinä käytetty data on kuitenkin jo käynyt läpi tarkistuksia ja korjauksia ennen kuin sitä on ladattu VIRTIA-opintotietopalveluun. VIRTIA-opintotietopalveluun ladattava sisältö on määrämuotoista, joten esimerkiksi puuttuvat tai mahdollisten asteikkojen ulkopuoliset arvot ovat tulleet ilmi ja korjattu, jotta data on saatu onnistuneesti ladattua palveluun.

Tässä tutkielmassa käytetty esimerkkidata on epätäydellistä, koska se sisältää useissa attribuuteissa pelkästään summa-arvoja. Datassa ei ole mukana alkuperäistä rivitietoa esimerkiksi siitä, mistä suorituksista tai ilmoittautumisista summat koostuvat. Tämä puutteellisuus rajaa mahdollisuuksia luoda sellaisia visualisointeja, joissa olisi mahdollista porautua yksittäisen opiskelijan opintosuorituskertymään tarkemmalla tasolla. Opintosuorituskertymään porautumalla voisi tarkastella esimerkiksi sitä, mitä suorituksia opiskelijalla on tai mitä häneltä on esimerkiksi tietyn lukuvuoden opinnoista jäänyt puuttumaan.

Opintosuoritusten raakadatasta eli suoraan perusrekisterin raporttikannasta tehty visualisointi mahdollistaisi edellä mainitun kaltaisen porautumisen. Se olisi samalla tämän prosessin vaiheen osalta merkittävästi monimutkaisempi toteutustapa, veisi merkittävästi enemmän aikaa ja olisi alttiimpi virheille. Suoraan raporttikannan päälle tehtynä visualisoinnista pystyttäisiin tekemään dynaaminen, eli se päivittyisi automaattisesti, kun siihen liittyvää dataa päivitetäisiin perusrekisteriin.

Tarkkojen tilastointi- ja laskentasääntöjen soveltaminen raakadataan vaatisi vahvan asiantuntemuksen lisäksi monimutkaisia kyselyjä, jotka yhdistelisivät oikealla tavalla useissa eri relaatiotietokannan tauluissa sijaitsevaa dataa. Tässä tutkielmassa tätä toteutusvaihtoehtoa ei tarkastella, koska perusrekisterin raporttikanta ei ole käytettävissä tutkielman tekohetkellä. Mielenkiintoinen jatkotutkimusaihe olisi myös vertailla vahvistettuja opetussuunnitelmia opiskelijoiden todellisiin suorituksiin. Näin pystyttäisiin tarkastelemaan, mistä opinnoista opiskelijoiden mahdolliset poikkeamat lukuvuosittaisessa opintopistemääräkertymässä yleisimmin johtuvat.

6.1. Datajoukon valinta

Visuaalisesti esitettävissä olevaa dataa on karkeasti lajiteltuna kolmenmuotoista: datan arvoja, datan rakennetta ja metadataa eli tietoa datasta. Datan arvojen esittämisessä valitaan jokin tai joitakin datan ominaisuuksia tai ulottuvuuksia, joita eri menetelmin vertaillaan suhteessa toisiinsa. Datan rakennetta visualisoitaessa keskitytään kuvaamaan datasta löytyviä keskinäisiä suhteita. Visualisoitavaa metadataa voivat olla esimerkiksi datan ominaisuuksien tilastolliset avainluvut tai datasta muulla tavoin koostetut arvot (esim. keskiarvot, mediaanit) tai datasta keinotekoisesti johdettu rakenne. [Tweedien, 1997] Esimerkkitapauksen tarkastelunäkökulma keskittyy datan arvojen ja datasta johdettujen arvojen vertailuun ja tarkasteluun, joten tämä täytyy ottaa huomioon myös datajoukon valinnassa.

Esimerkkidatassa on edellisessä vaiheessa määritellyn visualisoinnin tavoitteen näkökulmasta paljon ylimääräisiä attribuutteja. Ne rajataan pois valitsemalla datasta vain olennaiset attribuutit (kuva 16). Liian suppean datan visualisointi voi helposti johtaa käyttäjiä harhaan ja väärin johtopäätöksiin, mutta toisaalta liian suuren attribuuttijoukon valinta vaikeuttaa ja hidastaa visualisoinnin merkityksen sisäistämistä [Chittaro, 2006].

Attribuutti ”Ohjelma” sisältää tutkinto-ohjelman, jossa suoritettuja opintoja kyseessä olevan opiskelijan kohdalla tarkastellaan. ”Yht lv” on esimerkkivisualisoinnin näkökulmasta tärkein kenttä, sillä se kertoo opiskelijan edellisen lukuvuoden suoritusten yhteissumman opintopisteinä. ”Läsnäolokaudet” kertoo opiskelijan käyttämän opiskeluoikeuden määrän lukukausina. Se otetaan datajoukkoon mukaan, jotta pystyttäisiin tarkastelemaan opintopistemääräkertymää tarkemmin opintojen eri vaiheissa. ”Opiskeluoikeusavain” otetaan mukaan opiskelijan opiskeluoikeuden yksilöllisenä tunnisteena. Se ei ole suojattava henkilötieto, mutta sen avulla kohdealueen asiantuntijat pystyvät tarvittaessa yksilöimään ja jäljittämään yksittäisen opiskelijan datasta ja perusrekisteristä.

Dataa voidaan muuntaa esimerkiksi normalisoimalla sitä, luomalla käsitehierarkioita tai johtamalla olemassa olevasta datasta uusia ominaisuuksia [Han & Kamber, 2006]. Esimerkkitapauksessa alkuperäisen tavoitteenmäärittelyn takia on käytettävissä olevaa dataa rikastettava johtamalla siihen uusia arvoja. Loppukäyttäjien suosimaa hierarkiatasoa ”Yksikkö” ei raakadatasta löydy, joten se täytyy siihen lisätä. Yksiköllä tarkoitetaan tässä asiayhteydessä sitä organisaatiohierarkian osaa, jossa tietyn tutkinto-ohjelman opetus järjestetään. Kohdealueen organisaatorakenteeseen perustuen attribuutista ”Ohjelma” voidaan johtaa uusi korkeamman hierarkiatason attribuutti ”Yksikkö”. Tämäkin datan rikastaminen edellyttää kohdealueen tuntemusta, jotta pystytään oikeellisesti määrittelemään, mikä tutkinto-ohjelma kuuluu mihinkin yksikköön.

<input checked="" type="checkbox"/> Ohjelma
<input type="checkbox"/> Koulutustyyppi
<input type="checkbox"/> Koulutuksen opetuskieli
<input type="checkbox"/> Opiskelukunta
<input type="checkbox"/> Sukunimi
<input type="checkbox"/> Etunimet
<input type="checkbox"/> Ko. tutkintoon kirjoihintoajankohta ko. ammattikorkeakoulussa
<input type="checkbox"/> Kirjoillaolo ammattikorkeakoulussa 1.1.-31.7. (ilmoittautumistieto)
<input type="checkbox"/> Kirjoillaolo ammattikorkeakoulussa 20.9. (ilmoittautumistieto)
<input type="checkbox"/> Edellisenä syyslukukautena (1.8.-31.12.) suoritettujen varsinaisten opintopisteiden määrä
<input type="checkbox"/> Edellisenä kevätlukukautena (1.1.-31.7.) suoritettujen varsinaisten opintopisteiden määrä
<input checked="" type="checkbox"/> Yht. lv
<input checked="" type="checkbox"/> Läsnaololukukaudet
<input type="checkbox"/> Poissaololukukaudet
<input type="checkbox"/> Suoritettavien opintojen laajuus
<input type="checkbox"/> Ensimmäinen kirjoihintoajankohta ko. tutkintoon
<input type="checkbox"/> Lastentarhanopettajan pätevyyteen tähtäävät opinnot
<input type="checkbox"/> Rahoituslähde
<input type="checkbox"/> Maksullisen tilauskoulutuksen sijaintimaa
<input type="checkbox"/> Koulutuskoodi
<input type="checkbox"/> Opiskelija-avain
<input checked="" type="checkbox"/> Opiskeluoikeusavain

Kuva 16. Tarvittavien attribuuttien valinta datasta.

Datan muuntaminen on perusteltua myös attribuutin “Läsnaololukukaudet” kohdalla. Lukukausien lukumäärän hahmottaminen intuitiivisesti voi olla käyttäjille hankalampaa kuin käsitellä kulunutta opiskeluoikeusaikaa yleisemmin vuosina. Käytetty opiskeluaika saadaan laskettua systemaattisesti puolen vuoden tarkkuuteen jakamalla “Läsnaololukukaudet” kahdella. Tämä johdettu uusi attribuutti nimetään “Käytetty opiskeluaika” -muotoon.

Lisäksi yleisemmän tason visualisointien mahdollistamiseksi ja selkeyttämiseksi johdetaan vielä “Yht lv” –attribuutista uusi kaksiarvoinen pelkästään kyllä/ei-arvoja sisältävä attribuutti “Yli 55 op suorittanut”. Sillä pystytään helposti jaottelemaan opiskelijat kahteen alkuperäisen ongelmanmäärittelyn kannalta olennaiseen luokkaan.

Tässä tutkielmassa käytettyjen esimerkkien nominaaliarvot on anonymisoitu käyttämällä todellisten yksikköjen ja tutkinto-ohjelmien nimien sijaan keinotekoisia nimityksiä. Myös näytettävää

datajoukkoa on rajattu todellista pienemmäksi. Todellisissa käyttäjien saataville toimitetuissa prototyypeissä anonymisointia tai datajoukon supistamista ei ole tehty, vaan datan arvot ovat tunnistettavassa muodossaan ja lukumäärät todellisia.

6.2. Datatyypien määrittely

Valitun datajoukon attribuuttien datatyyppin määrittelemisen on visualisointiprosessissa tarpeellista. Tämä johtuu siitä, että tietyt visualisointimenetelmät sopivat erityisesti vain tiettyjen datatyyppien arvojen esittämiseen. Stevensin [1946] mukaan datan arvot voidaan jakaa neljään eri asteikkotyyppiin: nominaali-, ordinaali-, intervalli- ja suhdeasteikolle. Nominaaliasteikko toimii pelkästään tapausten ryhmittelyssä ja luokittelussa keskenään tasa-arvoisiin joukkoihin, eikä sillä pystytä mittaamaan datan ominaisuuden määrää. Esimerkkidatajoukon attribuutit “Ohjelma”, “Yksikkö”, “Yli 55 op suorittanut” ja “Opiskeluoikeusavain” ovat tätä datatyyppiä.

Ordinaaliasteikolla datan arvot kuuluvat luokkiin, joissa pelkästään luokkien keskinäisellä järjestyksellä on merkitystä, ei ryhmien välisillä keskinäisillä etäisyyksillä. Intervalliasteikossa sekä luokkien järjestys, että niiden väliset etäisyydet ovat merkityksellisiä. Suhdeasteikolla pystytään hyödyntämään tarkimmalla tasolla numeraalisten arvojen ilmaisuvoima, ja asteikolla on aina olemassa nollapiste [Ware, 2012]. Opintopistekertymää kuvaava “Yht. lv” sekä “Käytetty opiskeluaika” ovat luonteeltaan suhdeasteikon arvoja, mutta niitä voidaan tarvittaessa esittää karkeammalla tarkkuustasolla myös muilla asteikoilla.

Mackinlayn [1986] määrittelyssä datatyyppejä on vain kolme Stevensin esittämän neljän sijaan. Mackinlay tunnistaa määrällisen (quantitative) asteikon, joka vastaa parhaiten Stevensin suhdeasteikkoa, sekä ordinaali- ja nominaaliasteikot. Nowell et al. [2002] vähentävät Mackinlaynkin datatyyppejä vielä yhdellä. Heidän näkemyksensä mukaan ordinaaliasteikon data on käytännössä joko epätarkempi muoto määrällisestä datasta tai nominaalidatan erityinen järjestetty muoto.

Shneiderman [1996] on luonut datan tyyppien määrittelyyn toisenlaisesta näkökulmasta seitsemän kohdan taksonomian. Data voi Shneidermanin mukaan olla:

- yksiulotteista (1-dimensional), kuten tekstiä tai järjestettyjä listoja,
- kaksiulotteista (2-dimensional), kuten kartat,
- kolmiulotteista (3-dimensional), kuten 3D-grafiikka tai molekyylit,
- ajallista (temporal), jolloin tarkastelulla on yleensä alku ja loppupiste,
- moniulotteista (multi-dimensional), kuten tietovarastodata,
- puurakenteista (tree), kuten tiedostorakenne tai sukupuu,
- verkkorakenteista (network), kuten sosiaalisia verkostoja kuvaavat graavit.

Tämän luokittelun mukaan kaikki tässä tutkielmassa esimerkkinä käytettävä data on tyypiltään moniulotteista.

6.3. Datan puhdistaminen (data cleaning)

Datan esikäsittely sisältää usein monia erilaisia työvaiheita ja se on usein runsaasti aikaa vievä osuus visualisointiprosessissa. Sitä ei yleensä pystytä täysin automatisoimaan. Datan esikäsittely on koko prosessin kannalta erittäin tärkeä työvaihe, sillä huonolaatuisesta datasta ei voi syntyä kuin virheellisiä tai epäluotettavia visualisointeja. Korkealaatuisten visualisointien pohjalla on aina oltava totuudenmukaista dataa.

Dynaamiseen dataan pohjautuvissa visualisoinneissa on erityisen haasteellista datan laadun varmistaminen. Tällöin visualisointia varten tarvittava datan esikäsittelykin täytyisi pystyä suorittamaan automatisoidusti. Kaikkien mahdollisten puutteiden, virheiden ja muunnosten ennakointi ja toteuttaminen ohjelmallisesti voi olla haasteellista. Esimerkkivisualisointimme perustuu staattiseen tiedostoon, joten tässä tutkielmassa dynaamisen datan esikäsittelyn haasteisiin ei tarkemmin syvennytä. Esikäsittely on esimerkkiprosessissamme pystytty tekemään datalle kertaluonteisesti.

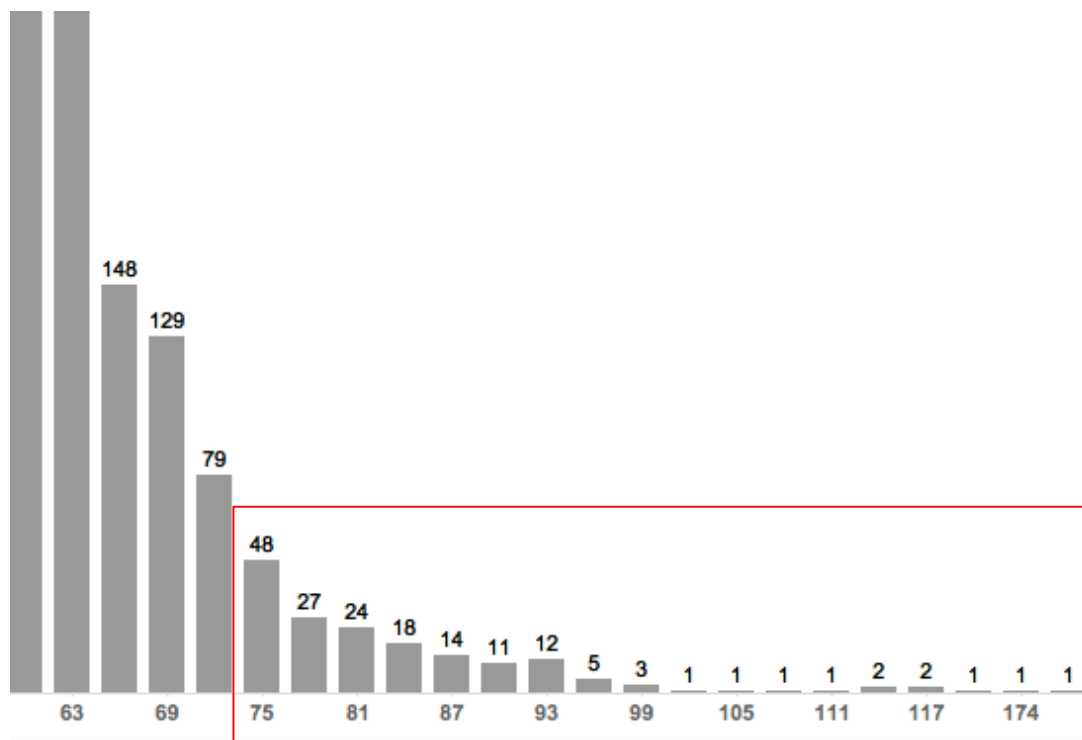
Dataa puhdistettaessa mahdolliset puuttuvat arvot pyritään ensisijaisesti täydentämään, jos se vain on mahdollista perustellulla työmäärällä ja käytettävissä olevalla asiantuntemuksella. Ellei täydennystä ole mahdollista tehdä ilman, että datan luotettavuus kärsii, yksi vaihtoehto on poistaa kokonaan tai korvata arvolla "tuntematon" puutteita sisältävät datan attribuutit tai tapaukset. Tämäkin on ongelmallista, mikäli puuttuvia arvoja on paljon tai ne kohdistuvat visualisoinnin kannalta merkittäviin dimensioihin tai suureen määrään tapauksia. Tuntemattomista arvoista voi syntyä vahingossa kiinnostavia hahmoja visualisointiin. Yksi menetelmä on generoida puuttuville arvoille täydennykset laskemalla keinotekoiset arvot muun datan pohjalta. [Tan et al., 2006]

Datan puhdistusvaiheessa myös samaa reaali maailman objektia kuvaavat moninkertaiset objektit eli duplikaatit pyritään poistamaan [González & Kobsa, 2003]. Esimerkkidatassamme puuttuvia arvoja tai samaa opiskeluoikeutta koskevia useita päällekkäisiä tapauksia ei ole, joten näitä toimenpiteitä ei esikäsittelyyn sisällynyt.

Kohina on datassa ilmeneviä satunnaisia virheitä, joita voi syntyä esimerkiksi väärinkäsitysten, inhimillisten tai tietojenkäsittelyssä tapahtuneiden virheiden seurauksena datan keräämisvaiheessa. Tällaiset virheet pyritään datan esikäsittelyssä joko korjaamaan, niiden merkitystä vähentämään tai poistamaan datasta kokonaan. Kohinan tunnistaminen ja korjaaminen voi olla haastavaa ja voi vaatia vahvaa datan kohdealueen tuntemusta esikäsittelijältä. Poikkeavia arvoja kohdellaan samoin kuin kohinaakin. Poikkeavat arvot voivat olla esimerkiksi huomattavan suuria tai pieniä muuhun dataan verrattuna tai olla jollakin muulla tapaa poikkeuksellisia. Poikkeuksellisten arvojen tapauksissa on toki mahdollista myös jättää käsiteltävät arvot osaksi datajoukkoa. [Han & Kamber, 2006]

Mahdollisia yksittäisiä poikkeavia arvoja asteikon ääripäistä voi haravoida helposti esimerkiksi järjestämällä datajoukon nousevaan tai laskevaan järjestykseen numeraalisia arvoja sisältävien kenttien osalta. Ääriarvoja voi etsiä myös visuaalisin keinoin esimerkiksi histogrammin avulla tutkimalla arvojen jakaumaa (kuva 17). Esimerkkiaineistosta löytyi useita tapauksia, joiden mukaan opiskelija olisi suorittanut edellisenä lukuvuonna yli 100 opintopistettä. Tämä käytännön realiteetit huomioiden ei ole kovinkaan todennäköistä. Kyseessä on todennäköisesti perusrekisteriin tehdyt virheelliset kirjaukset, jotka täytyy korjata.

Toisaalta poikkeavuuksia voi ilmetä myös tulkitsemalla usean eri attribuutin arvoja yhdessä [García et al., 2015]. Esimerkki tällaisesta tapauksesta voisi olla käytetyn opiskeluajan vertaaminen suoritettuun opintopistemäärään. Vain puoli vuotta opiskellut henkilö tuskin on ehtinyt kovinkaan suurella todennäköisyydellä suorittamaan 60 opintopistettä, vaikka se pidemmän ajan opiskelleelle olisikin aivan tavanomainen arvo.



Kuva 17. Otos opintopistekertymien jakaumaa kuvaavasta histogrammista poikkeavien arvojen tarkastelun apuna.

Esimerkkidatan opintopistemääräkertymien visuaalista jakaumaa tarkastelemalla jo esikäsittelyvaiheessa pystyi havaitsemaan nopeasti erään ilmiön. Sellaisten opiskelijoiden määrä, jotka eivät ole edellisenä lukuvuonna suorittaneet lainkaan opintoja, oli datassa melko suuri. Tämän havainnon pohjalta jatkokehityksessä voisi esimerkiksi ryhtyä tutkimaan tällaisia tapauksia tarkemmin. Tällöin visualisointiprosessi aloitettaisiin tästä lähtökohdasta alusta uusine

tavoitteenasetteluineen ja määrittelyineen. Visualisointiprosessin itseään suuntaava luonne näkyy erityisen hyvin tilanteissa, joissa alkuperäisiin tavoitteisiin pohjautuva visualisointi paljastaakin uuden kiinnostavan aihion uudelle visualisoinnille tai esimerkiksi muuttaa alkuperäistä tavoitetta tai määrittelyä tietämyksen lisääntyessä.

6.4. Operaatioiden määrittely

Interaktiivisuudella visualisoinneissa tarkoitetaan käyttäjän ja järjestelmän välillä tapahtuvaa vuorovaikutusta tutustuttamalla käyttäjä visualisoinnin avulla pohjalla olevaan datajoukkoon [Yi et al., 2007]. Lam [2008] määrittelee interaktiivisuuden joukkona toimintoja, jotka tekevät näkyviä muutoksia näytettävään visualisointiin. Visualisointiprosessin seuraavan vaiheen syötteeksi on määriteltävä operaatiot, jotka visualisoinnissa tulisi toteuttaa. Määrittelyn on perustuttava edellisessä vaiheessa tehtyyn tavoitteenasetteluun ja ongelmanmäärittelyyn, mutta pystyttävä kuvaamaan tehtäviä tietojenkäsittelylle tyypillisin termein kohdealueen erityissanaston ja -käsitteiden sijasta. [Munzner, 2009]

Esimerkkivisualisointimme tavoite on luoda uusia tarkastelunäkökulmia ja uutta ymmärrystä liittyen opiskelijoiden opintosuorituskertymään lukuvuoden aikana. Sen avulla kehityskohtiin pystyttäisiin paremmin puuttumaan ja näin lisäämään opiskelijoiden opintosuorituskertymää. Van Wijkin [2006 II] prosessimallin mukaan ratkaisumalleja kannattaa luoda useampia, jotta niistä pystyttäisiin valitsemaan tavoitteiden toteutumisen kannalta paras.

Interaktiivisuus ratkaisee yhden staattisille visualisoinneille tyypillisen perusongelman. Staattinen visualisointi pyrkii vastaamaan aina johonkin tai joihinkin ennalta asetettuihin kysymyksiin. Jokainen uusi kysymys vaatii käytännössä uuden visualisoinnin. Interaktiivinen visualisointi pystyy vastaamaan useampiin erilaisiin kysymyksiin, myös ennalta arvaamattomiin, sillä rajauksia ja näkökulmaa muuttamalla pystytään nostamaan datasta esiin kuhunkin kysymykseen liittyvä osa. [Tweedie, 1997]

Shneiderman [1996] on esittänyt visualisointioperaatioiden määrittelyn avuksi taksonomiaa, joka koostuu seitsemästä operaatiosta: yleisnäkymä (overview), zoomaus (zoom), suodatus (filter), lisäinformaatiota tarvittaessa (details-on-demand), suhteiden näyttäminen (relate), historia (history) ja poiminta (extract). Näistä korkean tason operaatioista valitaan kuhunkin visualisointiprosessiin soveltuvat ja niitä tarkennetaan konkreettisimmiksi.

Yleisnäkymäoperaatioilla halutaan antaa kokonaiskuva käsiteltävänä olevasta datasta. Tämä voidaan toteuttaa esimerkiksi näyttämällä jakaumia tai arvoja koko datasta kaikkein korkeimmalla abstraktiotasolla. Toisaalta yleisnäkymässä voidaan näyttää myös useista eri visualisoinneista koostettu kokonaisuus, jossa käyttäjälle tarjotaan mahdollisuus tarkastella itselleen kiinnostavinta osaa tarkemmin.

Zoomauksella voidaan laajasti ottaen käsittää käyttäjän mahdollisuutta tarkastella jotakin visualisoinnin osaa yksityiskohtaisemmin. Se ei välttämättä aina tarkoita zoomauksen yleiskielisen merkityksen tapaan kohteen suurentamista suhteessa muihin visuaalisiin elementteihin. Korostus voidaan tehdä myös muulla tapaa, kuten värien avulla, joko muuttaen valittua elementtiä tai häivyttämällä valitsematta jätettyjä elementtejä taustalle.

Suodatusoperaatiolla käyttäjä voi rajata visualisoinnin taustalla olevasta aineistosta pois esimerkiksi sellaisia luokitteluarvoja, jotka häntä eivät kiinnosta. Suodatusominaisuus on korkeakoulun kohdealueella tehdyissä visualisoinneissa erittäin tärkeä operaatio. Tarkastelunäkökulman rajaaminen yksikkö- tai tutkinto-ohjelmatasolle todettiin jo tavoitteenasettelussa tarpeelliseksi koskien kaikkia kohdealueen visualisointeja. Todellisia datan arvoja, joita jokin visuaalinen elementti kuvaa, tai niiden koostearvoja voidaan näyttää käyttäjille tarvittaessa lisäinformaationa. Se voi tapahtua esimerkiksi pienessä pop up -ikkunassa osoittamalla hiirellä kyseistä elementtiä.

Datassa ilmenevien suhteiden esittäminen on zoomauksen ja suodatuksen rinnalla yksi kaikkein keskeisimmistä operaatioista esimerkkiprojektissamme. Esimerkkidatajoukon visualisoinnissa voitaisiin yhtenä ratkaisuvaihtoehtona tarkastella opintopistekertymän suhdetta käytettyyn opiskeluaikaan. Tällöin nähdään, jakautuvatko 55 opintopistettä ylittävät kertymät tasaisesti koko opiskeluaikalle, vai onko havaittavissa, että esimerkiksi ensimmäisenä opiskeluvuonna 55 opintopistettä suorittaneita on muita opiskeluvuosia enemmän tai vähemmän. Näin ollen yksi operaatio, jonka visualisointi voisi toteuttaa, on paljastaa suhde attribuuttien “Käytetty opiskeluoikeusaika” ja “55 op suorittanut” välillä.

Historian säilyttäminen on operaatio, joka parantaa visualisoinnin käytettävyyttä. Se voidaan toteuttaa esimerkiksi mahdollistamalla siirtyminen saumattomasti edellisten ja seuraavien näkymien välillä kadottamatta informaatiota tai käyttäjän tekemiä valintoja. Poimintaoperaatiolla mahdollistetaan visualisoinneista valittujen alijoukkojen tai kyselyparametrien tallentaminen järjestelmästä jatkokäyttöä varten. Nämä operaatiot eivät esimerkkiprosessissamme yllä kovin korkealle prioriteetille. Pääpaino on operaatioilla suodatus, zoomauksen avulla toteutettava valinta sekä suhteiden esittäminen.

Edellä kuvattu Shneidermanin taksonomia ei ole ainut visualisointiin liittyvien operaatioiden kuvaamiseen luotu viitekehys, vaan niitä on kehitetty myös lukuisia muita. Luokittelut ovat osin varsin samankaltaisia, mutta osin myös poikkeavat toisistaan merkittävästikin sisällöltään tai tarkastelunäkökulmaltaan. Esimerkiksi Amar et al. [2005] ovat luoneet Shneidermanin osittain melko järjestelmätoimintokeskeisen taksonomian sijaan käyttäjälähtöisemmän kymmenen operaation taksonomian. Siinä painopiste on nimenomaan käyttäjien suorittamissa operaatioissa.

Tweedien [1997] mukaan vuorovaikutteiset toiminnot visualisoinneissa voidaan luokitella viiteen kategoriaan: piilottaminen ja suodatus (hiding/filtering), nimikoiminen tai kaksiarvoisen muuttujan lisääminen (esim. kyllä/ei) (labeling/Boolean encoding), animoitu navigointi visualisoinnissa (animated navigation), uudelleenjärjestely (reordering) ja algoritmin muunnos (algorithmic transformation).

Yi et al. [2007] puolestaan luokittelevat vuorovaikutteiset toiminnot käyttäjän tarkoitusperän mukaan. He eivät keskity siihen, millä tavoin järjestelmässä näiden toimintojen toteutus on tehty. Heidän määrittelemät interaktiivisuuden luokat ovat valinta (select), selailu (explore), uudelleenjärjestely (reconfigure), käytettävän visuaalisen vertauskuvan muuttaminen (encode), tarkastelutason muuttaminen (abstract/elaborate), suodatus (filter) ja yhdistäminen (connect). Tähän taksonomiaan on varsin helppo upottaa kaikkien yleisimpien tietovarastotyyppiseen dataan perustuvien visualisointien vuorovaikutustekniikat. Tosin suuri osa käytännön vuorovaikutustoiminnoista sopii useampaankin eri luokkaan samanaikaisesti, joten kiistattomasti erotteleva se ei ole.

Kuten edellä mainituista taksonomioista voidaan havaita, tunnistettuja visualisointeihin liittyviä vuorovaikutustapoja ja niiden luokitteluja on olemassa useita. On tarkoituksenmukaista harkita, mitä interaktiivisia toimintoja yksittäisessä visualisointiprosessissa asetettujen tavoitteiden saavuttamiseksi tarvitaan. Kaikkia toiminnallisuuksia ei kannata toteuttaa jokaiseen visualisointiin vain sen vuoksi, että se on mahdollista, ja tällä hetkellä vallalla oleva trendi. [van Wijk, 2006] Taksonomioiden tarkastelemista voi hyödyntää operaatioiden määrittelemisvaiheen apuna ikään kuin viitteellisinä luetteloina siitä, mikä kaikki voisi olla mahdollista. Erilaisten toimintojen käyttökelpoisuutta ja tarpeellisuutta kannattaa harkita oman visualisointiprosessin näkökulmasta ja pohtia, mitkä toiminnot parhaiten palvelisivat käyttäjää tavoitteiden saavuttamisessa.

Jos käyttäjän on mahdollista muokata visualisointia ja rajoituksia mielensä mukaan, tämä lisää väistämättä visualisoinnin subjektiivisuutta. Käyttäjä voi hienosäätää visualisoinnin sellaiseksi, että sen välittämä viesti tukee käyttäjän subjektiivista näkemystä asiasta, ja jättää jonkin toisen näkökulman huomiotta. Lisäksi interaktiivisuuden toteuttaminen lisää aina visualisoinnin toteutuskustannuksia sekä myös käyttäjien kognitiivista kuormaa, joten vuorovaikutteisia toimintoja tulisi käyttää vain perustellusti ja todelliseen tarpeeseen. [van Wijk, 2006]

Liialliset toiminnallisuudet toimivat samalla tavalla itseään vastaan kuin liiallisten visuaalisten elementtien ja tekstien käyttö. Niistä syntyy yhtä lailla visuaalista kohinaa, joka lopulta heikentää visualisoinnin kykyä välittää haluttua informaatiota. Erityisen ongelmallista on, mikäli toiminnallisuudet sisältävät osittaista päällekkäisyyttä [Niemi, 2015]. Määrittelyn tavoitteena kannattaakin pitää pienintä mahdollista joukkoa operaatioita, joilla tavoitteet vielä pystytään saavuttamaan, vaikkakin käyttäjien vaikutusmahdollisuudet jäisivätkin vähäisemmiksi [Lam, 2008].

7. Visuaalisen ratkaisun määrittely ja toteutus

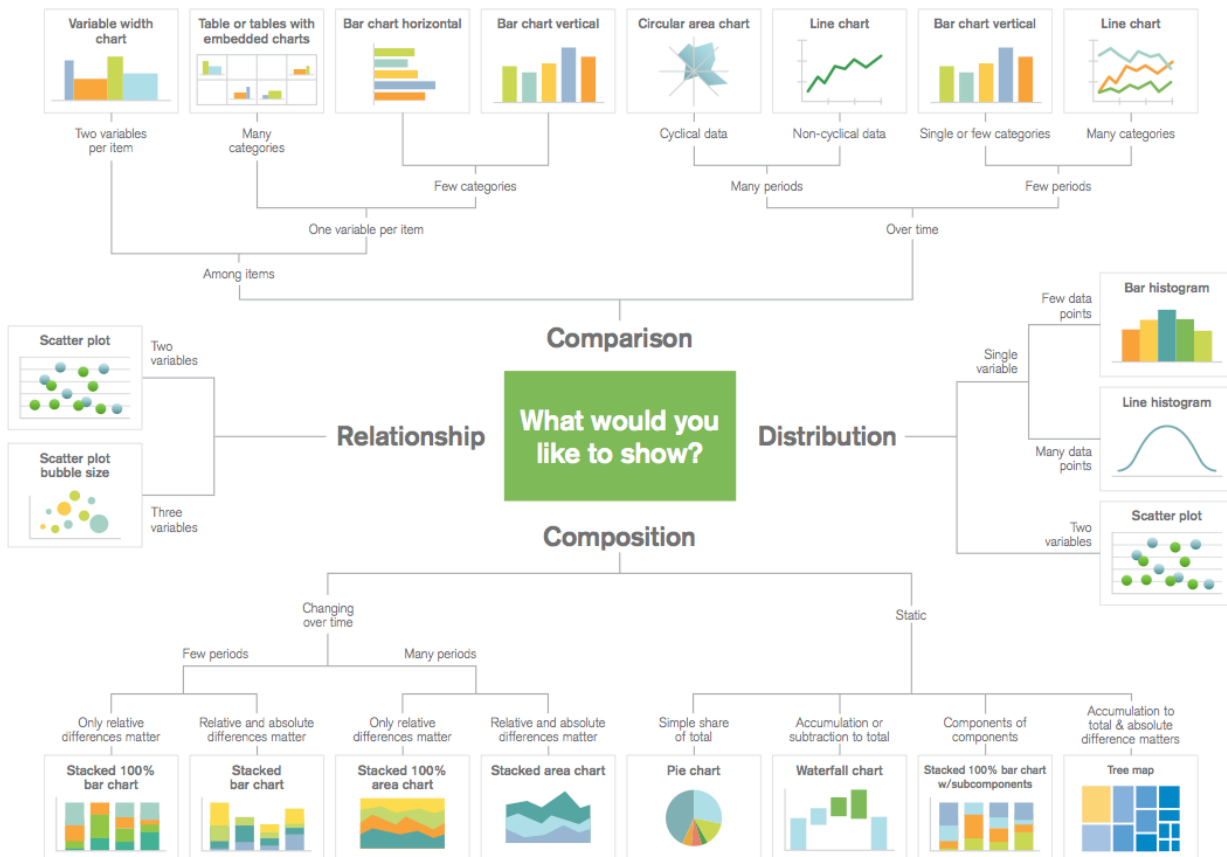
Visuaalisen ratkaisun määrittelyssä on hyödyllistä luoda käytettävissä olevien resurssien puitteissa useampia eri ratkaisuvaihtoehtoja ensimmäisessä vaiheessa määriteltyihin tavoitteisiin. Ensimmäinen mieleen tuleva ratkaisu ei välttämättä ole paras tai ainoa mahdollinen toteutustapa. [van Wijk, 2006 I] Valmiilla kaupallisilla visualisointisovelluksilla on suhteellisen helppo kokeilla erilaisia visualisointivaihtoehtoja ilman aikaa vievää koodaustyötä, vaikka lopullista implementaatiota ei kyseisillä kaupallisilla sovelluksilla lopulta tehtäisikään. Tässä tutkielmassa käyttäjien kokeiltavaksi tarkoitettujen prototyyppien tekemiseen on käytetty kaupallista Tableau -visualisointisovellusta (<http://www.tableau.com>).

7.1. Esitystavan valinta

Esitystavan valinnalla pyritään visuaalisin keinoin toteuttamaan prosessin ensimmäisessä vaiheessa määritellyt tavoitteet. Visualisointi kehitetään prosessin toisessa vaiheessa määrittelystä ja esikäsitellystä datasta sekä määritellyt operaatiot. Visualisoinnissa tasapainoillaan kahden keskenään ristiriitaisen pyrkimyksen välillä: esittää mahdollisimman paljon relevanttia informaatiota käyttäjälle ja samalla minimoida esitettävän informaation määrä visuaalisen selkeyden säilyttämiseksi [Roth & Mattis, 1990].

Esitystavan valintaa koskevan kirjallisuuden perusteella voidaan yhteenvedona todeta, ettei yksikään yksittäinen visuaalinen esitystapa ole sellainen, että se sopisi yhtä hyvin kaikille käyttäjäryhmille ja kaikissa mahdollisissa käyttötarkoituksissa. Tästä syystä visualisointiin ei ole olemassa suoraviivaista mallia tai luokittelua, jonka perusteella pystyisi helposti valitsemaan kuhunkin käyttötarkoitukseen parhaimman mahdollisen esitystavan. Jokaisessa käytännön prosessissa esitystavan valinta täytyy pohtia erikseen.

Näistä realiteeteista huolimatta A. Abela [2009] on luonut internetissä laajalle levinneen käytännönläheisen kaavion, jossa esitetään erityyppisiä visualisointiratkaisuja käytettäväksi erilaisiin tarpeisiin (kuva 18). Tietovarastotyyppisen datan visualisoimiseen yleisesti käytettävät esitystavat on kaaviossa jaettu neljään eri funktioon sillä perusteella, mitä visualisoinnilla tavoitellaan. Jokaiselle näkökulmalle on esitetty muutamia suositeltavia visualisointitapoja. Luokittelu ei ole täydellisen kattava, eivätkä esitetyt vaihtoehdot myöskään sulje toisiaan pois. Eri näkökulmia voi myös yhdistellä samaan esitystapaan tai samankaltaista kaaviota käyttää useammassa eri tarkoituksessa.



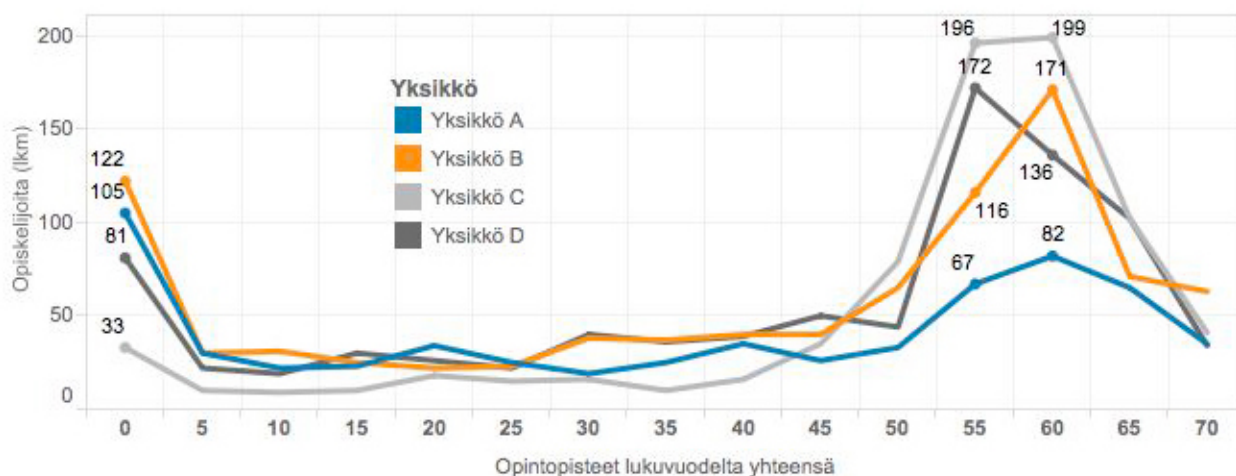
Kuva 18. Apuväline visuaalisen esitystavan valintaan [Abela, 2009].

Vertailulla (comparison) voidaan tarkastella rinnakkain esimerkiksi samankaltaista dataa eri ajanjaksoilta, vertailla datan arvoja keskenään tai eri organisaatioyksiköiden tuloksellisuuden mittareita keskenään. Vertailua voidaan visualisoida esimerkiksi erilaisilla pylväs- tai viivakaavioilla. Esimerkkivisualisoinnissamme ammattikorkeakoulun 55 opintopistettä suorittaneiden kokonaistarkastelun lisäksi loppukäyttäjät haluavat vertailla eri yksiköiden tai tutkinto-ohjelmien tuloksellisuutta tämän mittarin suhteen keskenään. Vertailun avulla pystyttäisiin esimerkiksi havaitsemaan, jos jossakin yksikössä tai tutkinto-ohjelmassa tulokset ovat keskimääräistä paremmat tai heikommat. Eroavaisuuksia havaitsemalla pystyttäisiin esimerkiksi levittämään hyviä käytänteitä tutkinto-ohjelmasta tai yksiköstä toiseen mm. opetussuunnittelussa, opiskelijoiden ohjauksessa tai muussa mahdollisessa eroavaisuuksia selittävässä tekijässä. Eroavaisuuksille voi olla olemassa myös muita selittäviä tekijöitä kuten erot tutkinto-ohjelmien vaativuudessa [Niemi, 2015].

Jakauman (distribution) esittämisellä pystytään kuvailemaan attribuutin erilaisten arvojen esiintymismäärää datassa. Jakaumaa visualisoidaan useimmiten histogrammeihin, pylväskaavioihin tai pistekaavioihin. Jakauman avulla pystytään näkemään, mitä arvoja datassa esiintyy keskimääräistä enemmän, sekä havaitsemaan poikkeamia ja mahdollisia ongelmakohtia.

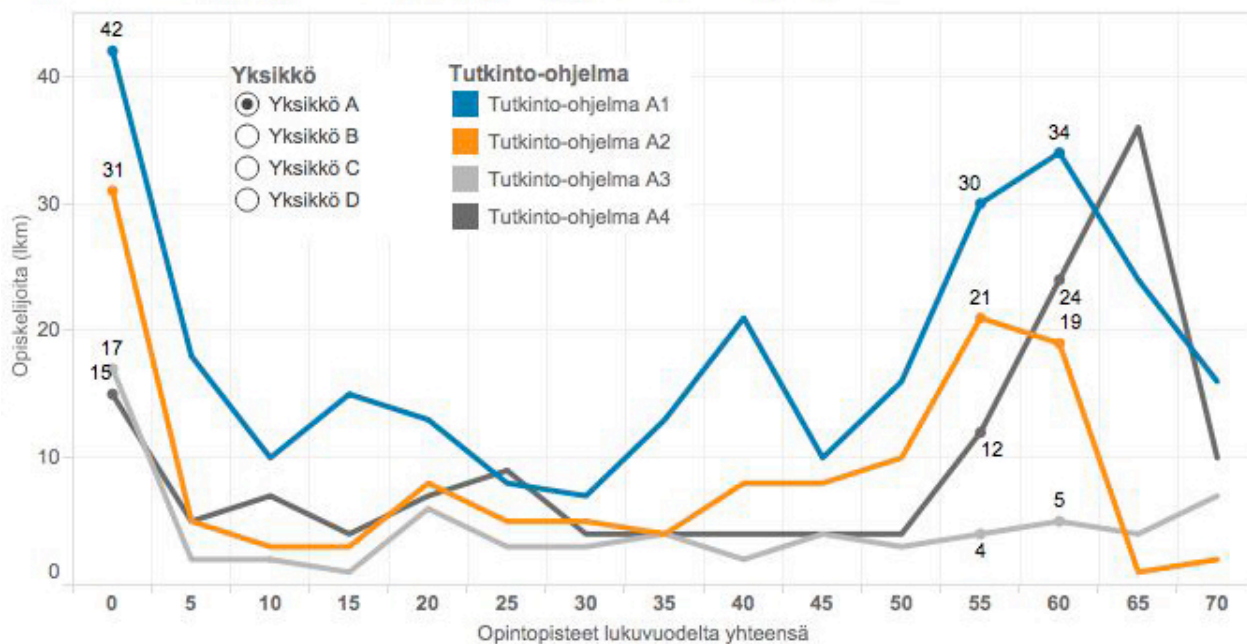
Jakaumia voi myös vertailla keskenään erilaisin luokitteluin, kuten esimerkkitapauksessamme tehtiin yksiköiden (kuva 19) sekä alemmalla hierarkiatasolla tietyn yksikön tutkinto-ohjelmien osalta (kuva 20). Vertailussa tiedostettu ongelma on kuitenkin kohdealueen yksiköiden ja tutkinto-ohjelmien erilaiset opiskelijavolyymit, joten jakaumaa tarkastellessa tämä seikka täytyy ottaa huomioon. Viivahistogrammista näkee tästä huolimatta nopeasti kunkin tarkastelukohteen jakauman huippukohtaan, eli terävimmän kärjen, joka ideaalitulanteessa olisi 55 tai 60 opintopisteen kohdassa.

Lukuvuoden opintopistekertymän jakauma yksiköittäin



Kuva 19. Lukuvuoden opintopistekertymän jakauma yksiköittäin absoluuttisina lukuina.

Lukuvuoden opintopistekertymän jakauma tutkinto-ohjelmittain

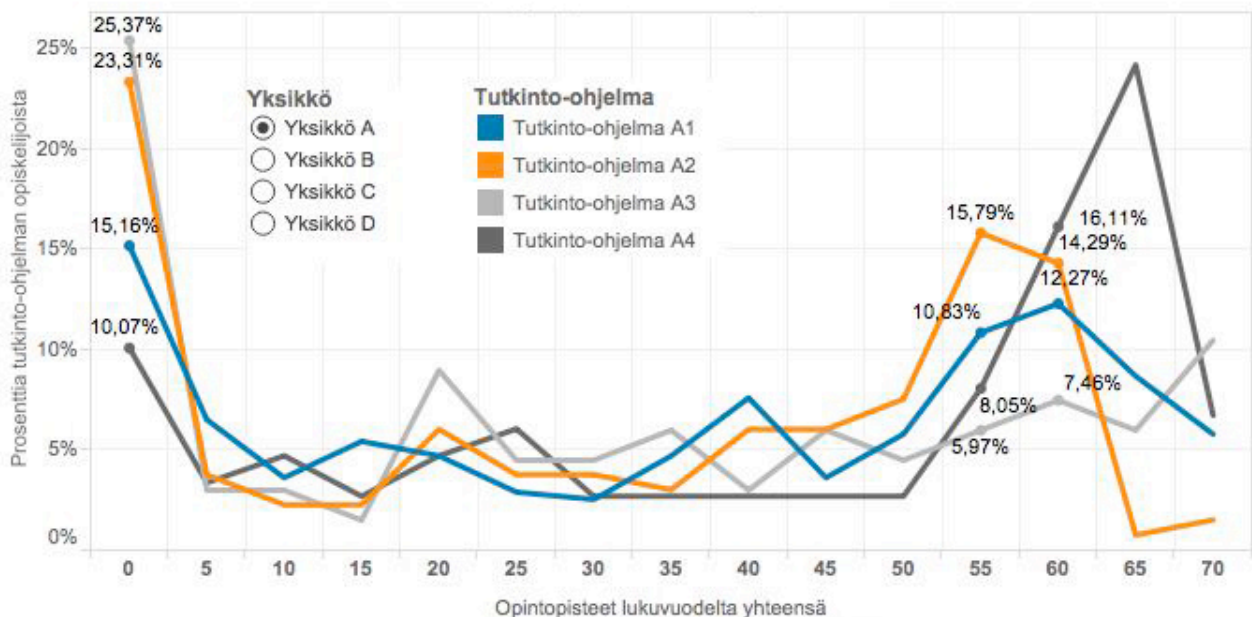


Kuva 20. Lukuvuoden opintopistekertymän jakauma tutkinto-ohjelmittain absoluuttisina lukuina.

Visualisointiprosessin ensimmäisen toteutuskierroksen päätteeksi tehdyn evaluoinnin jälkeen yksiköiden ja tutkinto-ohjelmien erilaisista opiskelijavolyymeista johtuvaan vertailun ongelmallisuuteen palattiin uudelleen seuraavassa iteraatiossa. Viivahistogrammeista kehitettiin uudet versiot, joissa vertaillaan tietyn opintopistemäärän suorittaneiden lukumäärien prosentuaalista suhdetta kunkin yksikön tai tutkinto-ohjelman opiskelijamäärään. Lopputulos osoittautui alkuperäistä absoluuttisiin lukumääriin perustuvaa jakaumaa merkittävästi vertailukelpoisemmaksi erityisesti tutkinto-ohjelmien tasolla (kuva 21).

Absoluuttisina lukuina tehdyssä vertailussa (kuva 20) Tutkinto-ohjelma A1 vaikutti suuremmasta opiskelijavolyymistaan johtuen tehokkaammalta 55 opintopisteen mittarin tuloksellisuuden näkökulmasta arvioiden kuin Tutkinto-ohjelma A2. Kun tulokset suhteutettiin kunkin tutkinto-ohjelman opiskelijamäärään, voidaan huomata, että todellisuudessa Tutkinto-ohjelma A2:n opiskelijoista suurempi osa on onnistunut A1:tä paremmin 55 opintopisteen kerryttämisessä lukuvuoden aikana. Lopulliseen implementaatioon onkin siis perustellumpaa tehdä vertailu suhteellisenä, koska vertailtavat luokat ovat volyymiltaan erikokoisia.

Lukuvuoden opintopistekertymän jakauma prosentteina tutkinto-ohjelman opiskelijamäärästä



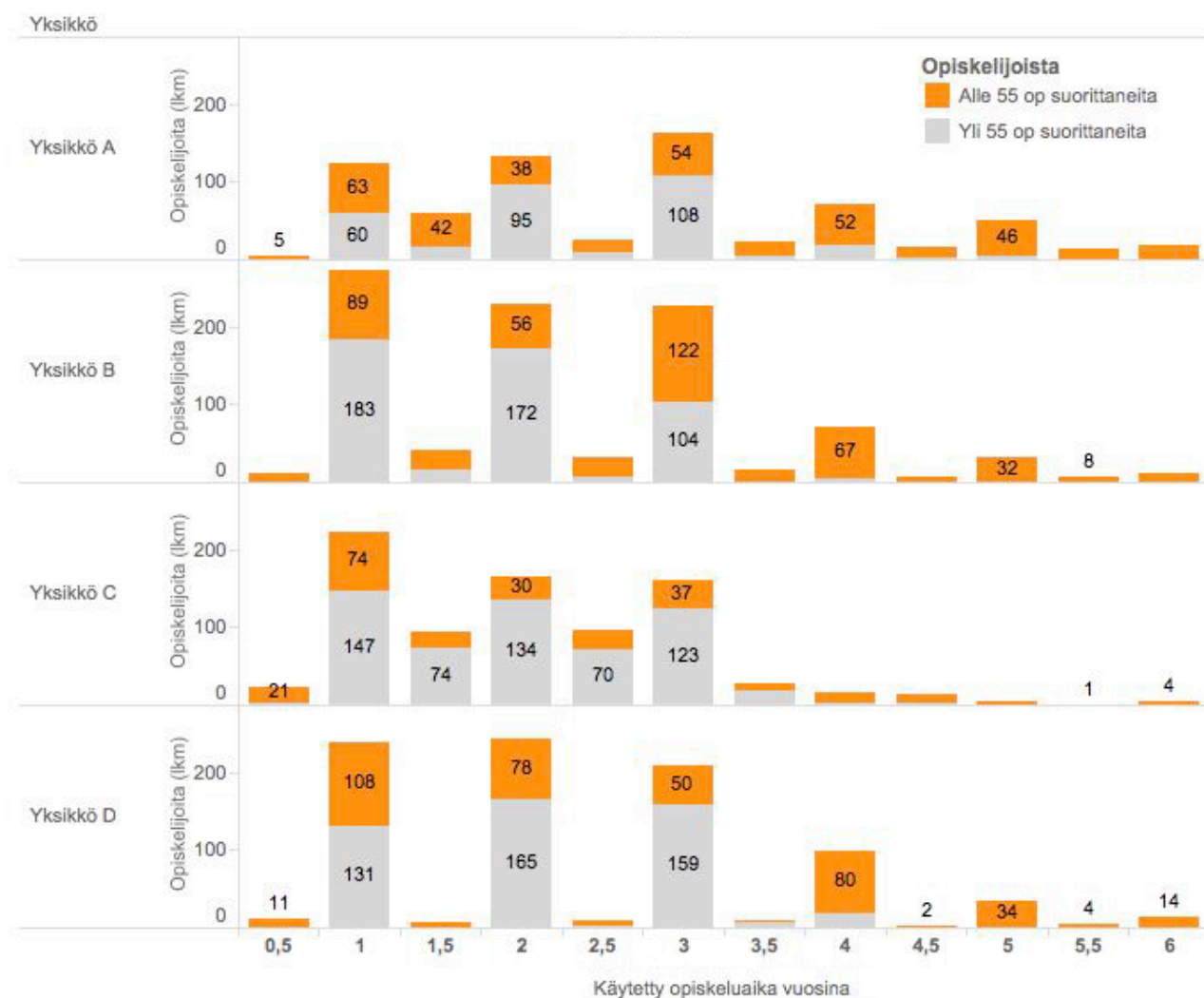
Kuva 21. Lukuvuoden opintopistekertymän jakauma prosentteina tutkinto-ohjelman opiskelijamäärästä

Koostumuksen (composition) visualisoinnissa päämääränä on esittää datan jonkun attribuutin määrän suhdetta kokonaisuuteen esimerkiksi piirakkakaavion (pie chart) avulla tai pinotuilla palkkikaavioilla (stacked bar chart). **Suhdetta** (relationship) visualisoimalla pystytään

tarkastelemaan esimerkiksi pistekaavion (scatter plot) avulla, millainen suhde kahden erillisen tekijän välillä on ja näin ollen päättämään, ovatko ne jollakin tavalla riippuvaisia toisistaan.

Esimerkkidatan koostumusta ja suhdetta voidaan visualisoida pinoamalla 55 opintopistettä suorittamattomat ja suorittaneet opiskelijat käytetyn opiskeluajan mukaan palkkikaavioksi (kuva 22). Näin dataan saadaan luotua näkökulma, josta pystytään havaitsemaan mahdollisia ongelmakohtia ja eroavaisuuksia käytettyjen opiskeluvuosien, yksiköiden tai tutkinto-ohjelmien välillä.

Yli ja alle 55 opintopistettä suorittaneet käytetyn opiskeluajan mukaan yksiköittäin



Kuva 22. Yli ja alle 55 opintopistettä suorittaneet käytetyn opiskeluajan mukaan yksiköittäin.

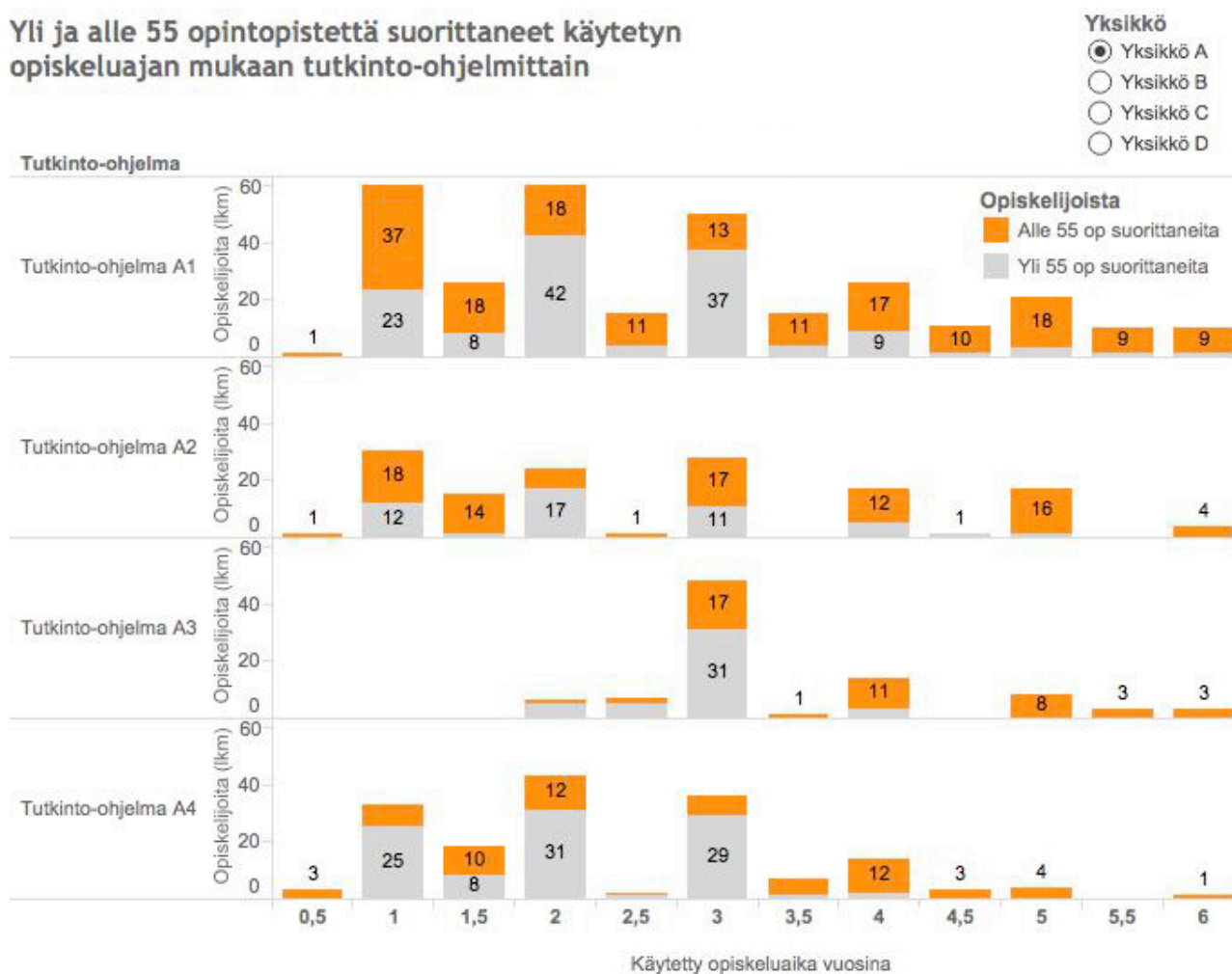
Tässäkin esimerkissä on yhdistelty käytännössä kolmea eri esitystapaa:

- **koostumusta** 55 opintopistettä suorittaneiden ja suorittamattomien välillä,
- **suhdetta** käytetyn opiskeluajan ja opintopistemääräkertymän välillä, sekä
- **vertailua** eri yksiköiden, käytettyjen opiskeluvuosien ja tutkinto-ohjelmien välillä.

Visualisoinnin avulla voidaan helposti havaita, että esimerkiksi Yksikkö A:n opiskelijoiden ensimmäistä lukuvuotta ja Yksikkö B:n kolmatta lukuvuotta kannattaisi tarkastella tarkemmin, sillä näissä suuri osa opiskelijoista jää alle 55 opintopisteen rajan.

Jotta voitaisiin saada vaivattomasti selville, mikä tai mitkä tietyn yksikön tutkinto-ohjelmista aiheuttavat tämän epätoivottavan tuloksen, visualisointi on syytä toteuttaa myös tutkinto-ohjelmatasolle yksiköittäin (kuva 23). Visualisoimalla myös alempi hierarkiataso saadaan selville, että Yksikön A tutkinto-ohjelmissa A1 ja A2 ensimmäisenä lukuvuonna opintopisteitä ei jostakin syystä kerry toivotulla tavalla.

Yli ja alle 55 opintopistettä suorittaneet käytetyn opiskeluajan mukaan tutkinto-ohjelmittain

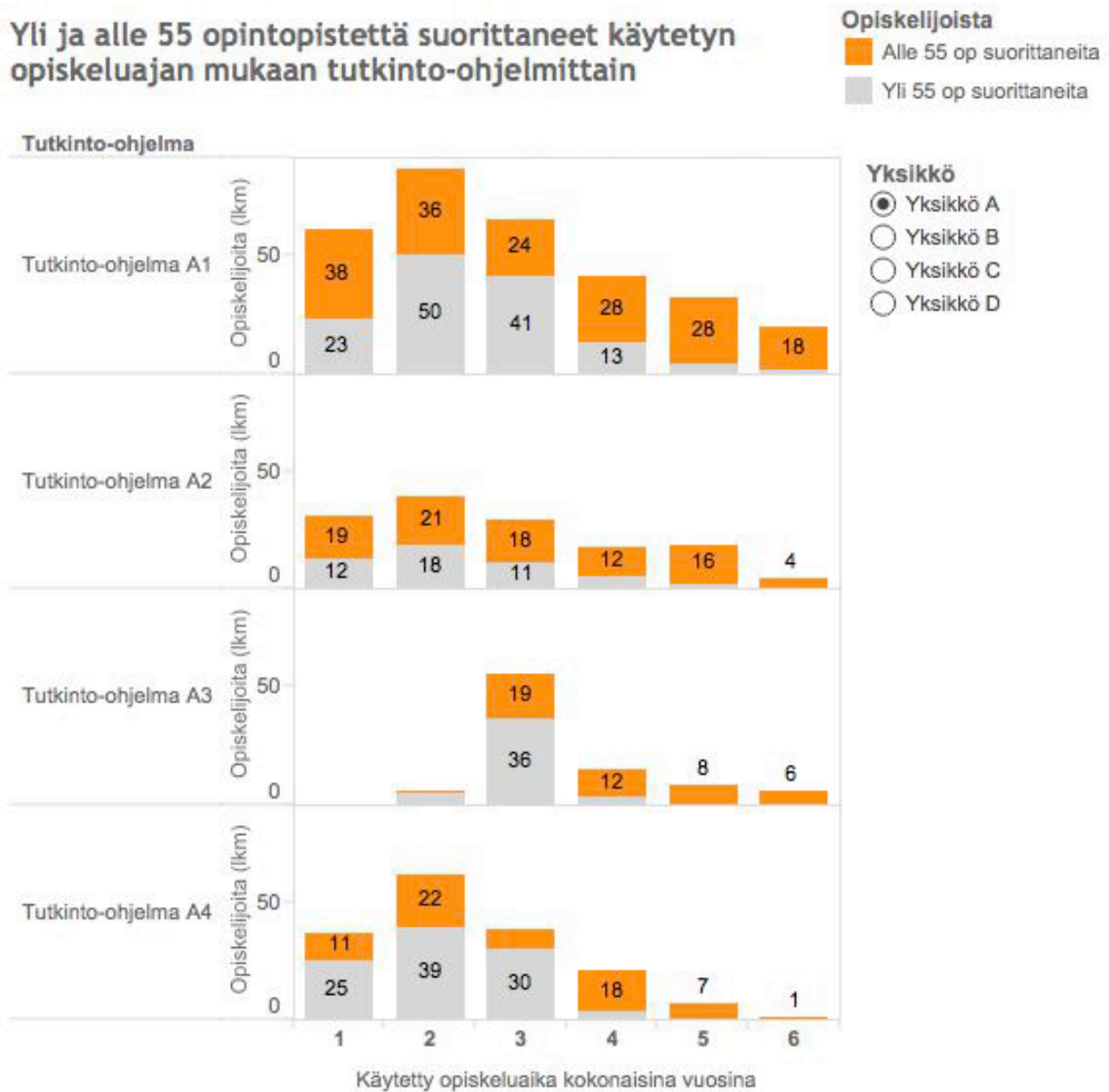


Kuva 23. Yli ja alle 55 opintopistettä suorittaneet käytetyn opiskeluajan mukaan tutkinto-ohjelmittain.

Visualisointiprosessin ensimmäisen läpikäynnin yhteydessä tehdyn evaluoinnin jälkeen pinottujen palkkien kaavioita yksinkertaistettiin pyöristämällä puolet vuodet kokonaisiksi. Tarkkuustasoa vähentämällä datan esikäsittelyssä päästiin selkeämpään ja intuitiivisemmin ymmärrettävempään

visuaaliseen lopputulokseen menettämättä kokonaiskuvan yksityiskohtaisuutta kohtuuttomasti (kuva 24).

Yli ja alle 55 opintopistettä suorittaneet käytetyn opiskeluajan mukaan tutkinto-ohjelmittain



Kuva 24. Yksinkertaistettu versio yli ja alle 55 opintopistettä suorittaneista tutkinto-ohjelmittain

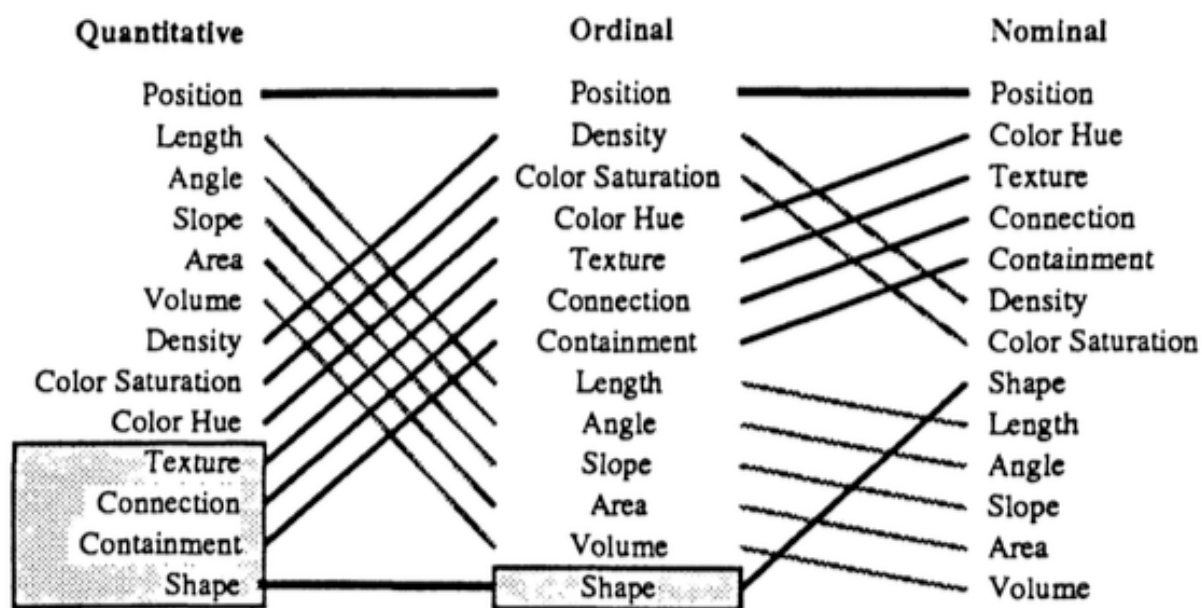
Viestin välittämisen kannalta tehokkainta on muodostaa graafisesta esityksestä sellainen, ettei se vaadi runsaasti selittävää tekstiä tai puhetta ollakseen ymmärrettävä. Sen pitäisi pystyä välittämään sanomansa myös itsessään. Ylimääräiset elementit toteutettavassa visualisoinnissa vähentävät esityksen tehokkuutta, koska ne johtavat lukijan huomiota muualle kuin kaikkein olennaisimpaan ja hidastavat varsinaisen viestin sisäistämistä [Tufte, 1983]. Visualisoinnin ymmärtämisen kannalta

olennaiset otsikot, selitteet ja muut merkinnät kannattaa aina asettaa lähelle niitä koskevaa visuaalista elementtiä käytettävyyden parantamiseksi [Schwabish, 2014].

Esimerkiksi yleisesti käytetyn datan visualisointityökalun Excelin valmiskaavioissa on pääsääntöisesti esillä suuri määrä tarpeettomia visuaalisia elementtejä tietyn visualisoitavan tehtävän näkökulmasta. Samoin niissä on toisinaan oletusarvoiset selitteet varsin kaukana varsinaisista dataa kuvaavista elementeistä. Tämän perusteella voisi väittää, että visualisointiin käytetty standardi ei välttämättä ole, ainakaan kirjallisuuden perusteella, lainkaan se paras tapa välittää informaatiota visuaalisesti. Siksi olisikin tärkeää saada muodostettua tiedon visualisoinnin soveltamiseen tieteellisiin faktoihin perustuvia standardeja, jolloin laajimmalle levinneisiin valmistyökaluihinkin voitaisiin saada oletukseksi parhaat mahdolliset visualisointivälineet.

7.2. Dataa kuvaavien visuaalisten elementtien valinta

Visualisoinneissa pystytään hyödyntämään monenlaisia eri keinoja kuvailemaan visualisoinnin pohjana olevaa dataa. Visualisoinnin keinoja ovat käytettävän visuaalisen elementin sijainti esityksessä (position), elementin värisävy (color hue), pintarakenne (texture), yhteys toiseen elementtiin (connection), sisältyvyys (containment), tiheys (density), värikylläisyys (color saturation), muoto (shape), pituus (length), kulma (angle), kaltevuus (slope), alue (area) ja volyyymi (volume). Mackinlay [1986] Hän on järjestänyt edellä luetellut elementtien visuaaliset ominaisuudet paremmuusjärjestykseen sen mukaan, miten tehokkaasti ihminen pystyy niiden avulla eri datatyyppejä havainnoimaan (kuva 25). Mackinlayn listaukseen viitataan kirjallisuudesta yhä paljon, vaikkakin sitä on myös kritisoitu paljon siksi, että sitä ei ole empiirisesti todennettu.



Kuva 25. Mackinlayn [1986] näkemys visuaalisten elementtien paremmuusjärjestyksessä erilaisten datatyyppeiden esittämisessä.

Erilaisten visuaalisten elementtien tehokkuuden luotettava empiirinen todentaminen on osoittautunut käytännössä hankalaksi. Visualisointien tavoitteet, käyttäjien suorittamat tehtävät sekä yksilölliset havainnointikyvyt vaihtelevat, jolloin laajemmin yleistettävissä oleva vertaileminen on ongelmallista toteuttaa. Laboratorioympäristössäkkin tällaisia yksilöllisiä muuttujia on vaikea hallita ja niiden merkitystä lopputulokseen aukottomasti selvittää. [Nowell et al., 2002]

Opintopistemääräkertymän jakaumaa kuvaavassa esimerkkivisualisoinnissa (kuvat 19-21) määrällisiä arvoja, eli tietyn opintopistemäärän suorittaneiden lukumäärää sekä suoritettuja opintopistemääriä, on kuvailtu hyödyntämällä viivoilla yhdistettyjen pisteiden sijaintia X- ja Y-asteikon rajaamalla alueella. Sijainti on Mackinlayn mukaan vahvin kuvaava tekijä visualisoinneissa kaikkien datatyypin yhteydessä.

7.3. Värien käyttö

Nominaaliasteikon mukainen datatyypin ”yksikkö” esimerkissämme on koodattu käyttämällä eri värejä eri yksiköiden visualisoinnissa. Koska sijaintikoodaus on jo käytössä muun ominaisuuden esittämisessä, täytyy erottuvuuden aikaansaamiseksi käyttää nominaaliasteikon kuvaamiseen jotakin muuta visuaalista ominaisuutta. Nowell et al. [2002] ovat toteuttaneet empiirisen tutkimuksen, jossa tarkasteltiin visuaalisten elementtien löytämisen nopeutta näyttöruudulta. Tutkimuksessa värien käyttö todettiin kaikkein tehokkaimmaksi visuaaliseksi keinoksi datatyypistä riippumatta silloin, kun tarkastellaan käyttäjien kykyä löytää tiettyjä visuaalisia elementtejä esityksestä. Myös Mackinlaynkin luokittelussa värisävy on määritelty nominaaliasteikon toiseksi parhaaksi vaihtoehdoksi, joten värien käyttöä yksiköiden erottelemiseksi voidaan pitää perusteltuna.

Värien käytöstä määrällisten arvojen kuvaamiseen on esitetty kirjallisuudessa paljon kritiikkiä. Esimerkiksi Kosslynin [1985] mukaan värit eivät luonnostaan muodosta luontaista jatkumoa esimerkiksi punaisesta vihreään, vaan lopputulos hahmotetaan helposti nominaaliryhmittelyinä, vaikka pyrkimys olisi kuvata määrällistä muutosta.

Visualisoinnissa käytettävien värien tulee riittävästi erottua toisistaan ja taustastaan, jotta niiden käyttö on tehokasta ja hyödyntää parhaalla mahdollisella tavalla ihmisen visuaalista havainnointikykyä. Tämä rajoittaa värien käyttöä tilanteissa, joissa erilaisia nominaaliarvoja on koodattavana paljon. Toisistaan selvästi erottuvia värejä on käytettävissä vain muutamia, koska visualisoinnin taustavärikin on otettava huomioon yhtenä käytössä olevana värinä. [Ware, 2012]

Kirjallisuudessa on toisistaan riittävästi erottuvien värien lukumäärää pyritty selvittämään jo vuosia. Yleisimmäksi käsitykseksi on muodostunut, että selkeästi toisistaan erottuvia perusvärejä on kuusi kappaletta. Tilanteissa, joissa värejä ei riitä kaikkien nominaaliarvojen erottelemiseen,

voidaan käyttää esimerkiksi muotoja, koska on olemassa monia selkeästi tunnistettavia ja toisistaan erottuvia muotoja. Liian samankaltaisten värien tai muotojen käyttäminen vaikeuttaa ja hidastaa aina visualisoinnin tulkitsemista. [Nowell et al., 2002] Värien valinnassa kannattaa huomioida myös soveltuvuus värisokeille käyttäjille [Ware, 2012], kuten kaikissa tämän tutkielman esimerkkivisualisointiprototyypeissa on tehty.

7.4. Selitteiden käyttö

Opintopistekertymää käytetyn opiskeluajan suhteen kuvaavassa visualisoinnissa (kuvat 22-24) nominaaliasteikon luokka “55 op suorittanut” on esitetty väreillä. Siinä ko. ominaisuuden eri arvojen esiintymien lukumäärää kuvaava määrällinen ulottuvuus on koodattu palkkien korkeudella sekä ilmaisemalla lukumäärät eksplisiittisesti palkkien yhteydessä. Selitteiden käyttäminen on tässä perusteltua, koska käyttäjä ei välttämättä pystyisi arvioimaan grafiikan esittämiä lukuarvoja kovin täsmällisesti ilman arvojen eksplisiittistä näyttämistä. Christin [1984] suorittamassa tutkimuksessa datan arvojen näyttäminen tuotti parhaimmat tulokset, kun tarkasteltiin, miten täsmällisesti testihenkilöt pystyivät tulkitsemaan eri visualisointitavoin esitettyä dataa.

Kaikkein matalimpien palkkien osalta arvoja ei kuitenkaan esimerkkivisualisoinnissa pystytä näyttämään ilman, että ne menisivät osin päällekkäin ja tekisivät visualisoinnista hankalasti luettavan. Pienen volyymin omaavilla palkeilla ei kuitenkaan tässä visualisoinnissa ole suurta merkitystä asetettujen tavoitteiden kannalta, joten niiden vähäinen huomioarvo ei muodostu ongelmalliseksi. Hyvin pienikokoisista visuaalisista elementeistä myös värin tai muodon erottamisen on osoitettu olevan hankalampaa, mikä on hyvä huomioida visualisointeja toteutettaessa [Nowell et al., 2002].

Jos visualisoinnille asetetun tavoitteen kannalta on tärkeää, että käyttäjien tulee saada määrällisen datan täsmällinen arvo tietoonsa, on selitteiden käyttäminen suositeltavaa, ellei se häiritse visualisoinnin selkeyttä olennaisesti. Useimmiten visualisoinneissa kuitenkin riittää, että käyttäjä pystyy hahmottamaan elementtien avulla arvojen suuruusluokan ja keskinäiset suuruussuhteet.

Otsikoiden ja selitteiden asettelussa kannattaa huomioida ihmisen tapa silmäillä visualisointeja. Otsikko on esimerkeissämme aina sijoitettu vasempaan ylälaitaan, koska käyttäjät silmäilevät visualisointeja yleensä samansuuntaisesti kuin lukevat. [Harvard, 2014] Esimerkkitapauksessamme potentiaaliset käyttäjät ovat suomalaisia, joten heidän lukusuuntansa on vasemmalta oikealle ja ylhäältä alas. Värikoodausta selittävät tekstit pyritään sijoittamaan lähelle visuaalisia elementtejä, jotta ne ovat helposti löydettävissä tulkinnan avuksi [Harvard, 2014].

7.5. Operaatioiden toteutus

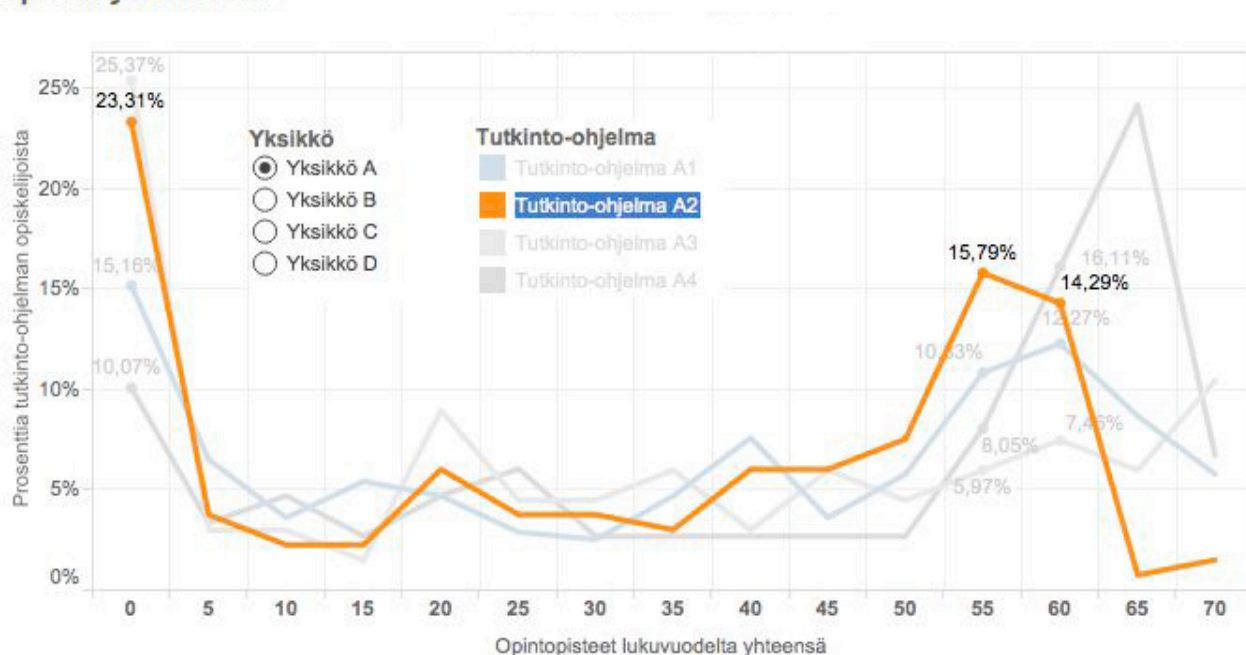
Visualisoinnin interaktiivisuus saadaan aikaan toteuttamalla prosessin toisessa päävaiheessa määritellyt operaatiot. Ilman käyttäjälle tarjottavia aktiivisia vuorovaikutusmekanismeja

visualisointi on käytännössä pelkkä staattinen esitys tai joukko staattisia esityksiä, joiden sisältöön käyttäjä ei pysty vaikuttamaan. Staattisten esitysten käyttö visualisoinnissa voi olla toisinaan täysin perusteltua, mikäli datajoukko on riittävän pieni. Toisin sanoen siinä on tällöin vain vähän visualisoinnin kannalta olennaisia attribuutteja tai sitä halutaan käyttää esimerkiksi tietyn ennalta havaitun ilmiön esittämiseen. [Yi et al., 2007]

Staattiset visualisoinnit eivät samalla tavalla kuormita käyttäjiä kognitiivisesti kuin monipuolisilla vaikuttamismahdollisuuksilla varustetut visualisoinnit, koska käyttäjiltä ei vaadita erityisiä toimenpiteitä halutun näkymän aikaansaamiseksi. Toisaalta interaktiivinen visualisointi usein vie fyysisesti yhdellä kertaa vähemmän tilaa kuin samaan lopputulokseen tähtäävä joukko staattisia visualisointeja, joita usein jouduttaisiin esittämään useita rinnakkain. Visualisointiprosessissa onkin tärkeää löytää tavoitteiden kannalta sopiva tasapaino interaktiivisuuden ja staattisuuden välillä. [Yi et al., 2007]

Vuorovaikuttaisia toimintoja toteuttamalla visualisointiin voidaan liittää käyttäjille mahdollisuus selvittää jonkin yksittäisen elementin arvo esim. zoomaamalla sitä tai viemällä osoittimen sen päälle, jolloin arvo näytetään. Näin pystytään vähentämään visuaalisesti kuormittavien pysyvien selitteiden määrää. Tämän tutkielman visualisointiprototyypeissä täsmällisten arvojen näyttäminen toteutettiin staattisten selitteiden lisäksi pienillä pop up –ikkunoilla, jotka avautuvat hiiren osoittimen ollessa jonkin visuaalisen elementin päällä ja näyttävät tarkasteltavana olevien tekijöiden arvot kyseisessä kohdassa.

Lukuvuoden opintopistekertymän jakauma prosentteina tutkinto-ohjelman opiskelijamäärästä



Kuva 26. Esimerkki valintaoperaation käytöstä.

Valintaoperaatioissa klikkaamalla jotakin visuaalista elementtiä muut elementit himmenevät taustalle ja valittu elementti korostuu ruudulla (kuva 26). Esimerkiksi viivahistogrammina toteutetussa prototyypissämme päällekkäin kasautuvia viivoja on joissain kohdin varsin vaikea erottaa toisistaan. Tällaisessa visualisoinnissa valintaoperaation avulla käyttäjä pystyy klikkaamalla helposti valitsemaan tarkasteltavaksi juuri sen tai ne visuaaliset elementit, joista hän on tietyllä hetkellä kiinnostunut menettämättä kuitenkaan ruudulta kokonaan kontekstina toimivia muita visuaalisia elementtejä.

Suhteen näyttämiseen tarvittava operaatio toteutettiin prototyypeissä staattisena ja ennalta määriteltynä ominaisuutena. Arvot on kiinnitetty X- ja Y-asteikoille pysyvästi, eikä käyttäjä pysty niitä muuttamaan. Käytettävässä datajoukossa oli merkityksellisiä attribuutteja sen verran vähän, että suhteen olennainen informaatio pystyttiin esittämään staattisesti. Sen sijaan suodatusmahdollisuus on annettu kaikissa visualisointiprototyypeissämme käyttäjien saataville, koska silloin he voivat tarkastella datajoukkoa haluamaltaan hierarkia- tai luokkatasolta. Prototyypeissä on pyritty yksinkertaisuuteen interaktiivisuuden osalta. Jatkokehityksessä toimintoja voidaan monipuolistaa, mikäli käyttäjät kokevat sen tarpeelliseksi ja lisätoimintojen tarpeellisuus pystytään perustelemaan.

Prototyypeihin mahdollistettiin erityisesti datan puhtaudesta ja yhteneväisyydestä vastaavia opintoasiainhallinnon asiantuntijoita varten operaatio, jonka avulla voidaan tarkastella visualisoinnin taustalla olevaa raakadatan relevanttia osajoukkoa. Tätä operaatiota voidaan käyttää myös apuna visualisoinnin sisällön validoimisessa. Esimerkiksi klikkaamalla 0 opintopistettä suorittaneiden kohdalla hiiren oikeaa painiketta voi ruudulle generoitavasta uudesta ikkunasta tarkastella niitä raakadatan opiskelijarivejä, joista visuaalinen elementti on koostunut.

8. Evaluointi

Visualisointien evaluoinnissa on hyödynnetty perinteisesti paljon HCI:n alueella kehitettyjä ja yleisesti käytettyjä evaluointimenetelmiä. Näitä on kehitetty erityisesti käyttöliittymien ja käytettävyyden evaluointia varten. Viime vuosina on kuitenkin kirjallisuudessa alkanut esiintyä enenevässä määrin myös erityisesti visualisointien evaluointiin keskittyviä tutkimuksia. [Chittaro, 2006]

Munznerin [2009] nelitasoisen prosessimallin mukaan evaluointi ei ole pelkästään yksittäinen erillinen vaihe prosessin lopussa. Validointia on tärkeä tehdä kaikissa prosessin eri vaiheissa, jotta tasolta toiselle periytyviltä virheilta vältyttäisiin. Munzner käyttääkin evaluoinnin rinnalla termejä validointi ja verifiointi, joiden tarkoituksena on korostaa evaluoinnissa kahta näkökulmaa. Toisaalta visualisointiprosessin tuotoksena toteutetun tuotteen pitää olla vastaus oikeaan ongelmaan ja toisaalta sen toteutus pitää olla oikeellinen, eli sen tulee toimia tarkoituksenmukaisella tavalla. Evaluoinnin tulisi olla elimellinen osa suunnitteluprosessia, jotta perustavanlaatuisilta virheilta vältyttäisiin.

Koska tästä tutkielmasta on visualisoinnin varsinainen implementaatio rajattu pois, myöskään evaluointia ei pystytä koko laajuudessaan soveltamaan esimerkkiprosessiimme. Määrittelyvaiheen päätelmien oikeellisuutta sekä prototyyppien ymmärrettävyyttä ja validiutta on kuitenkin arvioitu kohdealueen asiantuntijaedustajien kanssa prosessin eri vaiheissa.

8.1. Vaiheittainen evaluointi

Prosessin ensimmäisen vaiheen jälkeen suunnittelijalla pitäisi olla täsmällisesti tiedossa kohdealueen ongelmat, joihin visualisoinnin keinoin pyritään löytämään ratkaisua. Jos vaihe ei toteudu tarkoituksenmukaisella tavalla, suurin uhka on se, että prosessin seuraavissa vaiheissa ryhdytään luomaan visualisointiratkaisua sellaiseen ongelmaan, jota käyttäjillä ei edes ole. Tästä syystä ensimmäisen vaiheen lopuksi suunnittelijan olisikin validoitava muodostamansa käsitys kohdealueen ongelmasta varsinaisilla käyttäjillä laadullisen tutkimuksen menetelmin esimerkiksi käyttämällä haastatteluja ja havainnointia. [Munzner, 2009]

Kohdealueen ongelmanmäärittelyä voidaan pitää käyttäjiltä saadun palautteen perusteella validina, sillä 55 opintopistettä suorittaneiden määrän lisääminen on kaikkien potentiaalisten käyttäjäryhmien yhteinen tavoite. Vasta visualisoinnin varsinaisen implementaation käyttöönottoaste kuitenkin osoittaa sen, koetaanko visualisointi menetelmänä hyödylliseksi apuvälineeksi tämän tavoitteen saavuttamisessa. Mahdollista on, että päämäärään pyritään edelleen mieluummin joillakin muilla aiemmin tutuksi tulleilla keinoilla.

Toisen vaiheen suurimpana uhkana voidaan pitää sitä, että valittu data ja määritellyt operaatiot eivät mahdollista ensimmäisessä vaiheessa määriteltyjen ongelmien ratkaisua. Varmin tapa varmistaa, onko tuotettu visualisointi tarkoituksenmukainen tunnistettuun tarpeeseen, on pyytää kohdealueen käyttäjiä tarkastelemaan prototyyppiä todellisten työkulkujensa yhteydessä. Visualisointisovelluksen suunnittelijan ei kannata keksiä evaluoijille keinotekoisia testaustehtäviä irrallaan todellisuudesta. Näin myös suunnittelija saa mahdollisuuden nähdä, miten visualisointia käytetään sovellusympäristössä ja miten todelliset käyttäjät ymmärtävät suunnitteluprosessin kuluessa tehdyt valinnat. [Munzner, 2009]

Toisen vaiheen tavoitteiden evaluointi osoittautui hieman ongelmalliseksi esimerkkiprosessissamme. Tämä johtui päätöksestä käsitellä pääasiallisena käyttäjäryhmänä opintoasiainhallinnon asiantuntijoita. He hallinnoivat visualisoinnissa käytettävää dataa, koordinoivat sen keräämistä ja kirjaustapoja, sekä välittävät ja tiedottavat siitä muille toimijoille. Opintoasiainhallinnon asiantuntijat eivät antamassaan palautteessa niinkään kommentoineet visualisointien mahdollisia hyötyjä omien työtehtäviensä ja prosessiensa kehittämisen näkökulmasta, vaan viittasivat lähinnä muiden mahdollisten käyttäjien visualisoinneista saamaan hyötyyn. Jatkokehityksessä opintoasiainhallinnon rooli suhteessa muihin potentiaalisiin käyttäjiin täytyisikin ottaa tarkempaan selvitykseen, jotta potentiaalisen käyttäjäryhmän rajaus olisi yksiselitteisempi.

Toisen vaiheen evaluointi esimerkkiprosessissamme oli myös siitä näkökulmasta hieman poikkeuksellinen, että visualisoinnit tehtiin lähtökohtaisesti tiettyyn rajattuun, ennalta käsiteltyyn dataan perustuen. Tästä johtuen toisessa vaiheessa ei datajoukkoa valittu ja rajattu samanlaisesta vaihtoehtojen joukosta kuin sellaisessa visualisointiprosessissa, jossa lähdettäisiin puhtaasti tavoitteiden pohjalta miettimään, mitä dataa mistäkin lähteistä hyödynnettäisiin. Visualisoinneissa oli lähtökohtana tietty varsin suppea datajoukko, josta pyrittiin parhaalla mahdollisella tavalla löytämään uusia näkökulmia aihealueen tarkastelemiseen. Jos käytettävissä olisi ollut kaikki aihepiiriä sivuavat olemassa olevat tietolähteet ammattikorkeakoulussa, niin prosessin toisessa vaiheessa olisi jouduttu tekemään isompia ratkaisuja ja suurempi esikäsittelytyö. Tällöin myös evaluoinnin merkitys olisi ollut suurempi tehtyjen ratkaisujen oikeellisuuden arvioinnissa.

Kolmannen vaiheen onnistumista voidaan arvioida sillä, miten hyvin käytetyt visuaaliset metaforat välittyvät käyttäjille ja miten hyvin he pystyvät niiden avulla informaatiota sisäistämään ja visualisointeja tulkitsemaan. Kolmannen vaiheen onnistumista voidaan arvioida myös analyttisesti esim. tutkimalla, miten visualisoinnissa on huomioitu tunnettuja sisäistämiseen ja havainnointikykyyn liittyviä periaatteita ja miten hyvin ne noudattavat yleisiä heuristisia sääntöjä. Eräs perinteinen evaluointimenetelmä on järjestää visualisoinnille muodollinen käyttäjätkimys laboratorio-olosuhteissa. Näin voidaan tutkia esimerkiksi, miten kauan visualisoinnin merkityksen hahmottamiseen tai jonkin tiedon löytämiseen kuluu aikaa. Tutkia voidaan myös sitä, kuinka paljon

ja minkälaisia virheitä graafisten esitysten tulkitsemisessa syntyy. Valittujen algoritmien ja toteutustavan tehokkuutta voidaan myös arvioida tutkimalla suorituskykyä ajan ja tarvittavan muistin suhteen. [Munzner, 2009]

8.2. Kokonaisevaluointi

Visualisoinnin valmistuttua voidaan arvioida kaikkien vaiheiden onnistumista tarkastelemalla luodun visualisoinnin hyväksikäytön määrää sekä tarkkailemalla, miten käyttäjät arjessaan visualisointia hyödyntävät [Munzner, 2009]. Evaluointi voidaan suorittaa joko kohdealueen käyttäjien toimesta, jolloin menetelmää voidaan kutsua testaukseksi (testing method) tai käytettävyyteen erikoistuneiden ammattilaisten toimesta, jolloin voidaan puhua tarkastuksesta (inspection method) [Andrews, 2008].

Komlodi et al. [2004] ryhmittelivät kirjallisuudessa esitetyt tiedon visualisoinnin käyttäjätutkimukset neljään teemaluokkaan sen mukaan, mistä näkökulmasta evaluointia oli tehty. Kontrolloiduissa kokeissa vertaillaan erilaisia visuaalisia elementtejä tai visualisoinnin muissa yksityiskohdissa tehtyjä valintoja keskenään (controlled experiments comparing design elements). Esimerkkimme visualisointiprosessissa tällainen kokeilu voisi olla yksinkertaisimmillaan esimerkiksi sen vertaileminen, toimiiko tarkasteltavan yksikön valitsemisessa käyttäjien mielestä paremmin pudotusvalikko vai radionappivalikko. Tämänkaltaista evaluointia ei kuitenkaan tutkielman yhteydessä pidetty tarpeellisena toteuttaa, koska sitä ei pidetty olennaisena visualisointiprosessin tutkimisen näkökulmasta.

Käytettävyytutkimuksissa (usability evaluation) pyritään löytämään käyttäjien visualisoinnissa kohtaamia ongelmia, jotta ne voitaisiin korjata seuraavissa iteraatioissa tai jatkokehityksessä [Komlodi et al., 2004]. Käytettävyytutkimus on esimerkkiprosessissamme hyödyllinen erityisesti siinä vaiheessa, kun visualisointi on toteutettu kokonaisuudessaan laajemman käyttäjäjoukon saataville. Tällöin prototyypivaiheiden mahdolliset käytettävyysongelmat voitaisiin havaita ja korjata. Tämä lähestymistapa on visualisoinnin kehittäjän kannalta tehokas ja mielekäs, sillä se tuottaa täsmällistä tietoa korjattavista osa-alueista jatkokehityksen tueksi. Potentiaalisille käyttäjille tällainen testaus ei kuitenkaan ole kovin mielekästä, jos testattavassa versiossa on vielä paljon virheitä tai jos testaamiseen käytetään keinotekoisia tehtäviä. [Plaisant, 2004]

Erilaisten visualisointityökalujen vertailuun tähtäävissä kontrolloiduissa kokeissa (controlled experiments comparing two or more tools) käytetään samassa kontekstissa samaan tarkoitukseen kahta tai useampaa erilaista visualisointimenetelmää tai -työkalua. Kokeissa vertaillaan niiden toivottuja ja ei-toivottuja piirteitä toisiinsa. Tällaisesta evaluoinnista olisi hyötyä esimerkiksi tilanteessa, jossa harkitaan ohjelmiston ostamista visualisointisovellutusten tekemiseen organisaatiolle. Tällä menetelmällä voidaan vertailla eri järjestelmien käyttökelpoisuutta kyseisen organisaation toimintaympäristössä ja kohdealueen datan esittämisessä.

Todellisessa toimintaympäristössä tehdyt käyttäjätutkimukset (case studies of tools in realistic settings) ovat usein toteutetuista evaluointimenetelmistä kaikkein harvemmin käytettyjä. Niiden toteutus vie paljon aikaa, eikä tuloksetkaan ole helposti toistettavissa tai yleistettävissä. Tästä syystä tällaiset käyttäjätutkimukset eivät ole suuresti houkutelleet tutkijoita. Heidän intressinsä on yleensä tuottaa yleistettävissä olevia ratkaisuja ja tietämystä, jota voitaisiin soveltaa laajemminkin kohdealueesta tai yksittäisen visualisointiprosessin yksilöllisistä piirteistä riippumattomasti. [Komlodi et al., 2004]

Yksittäisen kohdealueen kannalta tällaisilla tutkimuksilla on suuri käytännön merkitys. Vain näin pystytään arvioimaan visualisoinnin merkitystä ja vaikuttavuutta todellisessa toimintaympäristössä ja todellisten käyttäjien keskuudessa eli osana todellisia työnkuluja. [Komlodi et al., 2004] Todelliseen toimintaympäristöön sijoittuva evaluointi on tässä tutkielmassa esitetyn esimerkkivisualisointiprosessin kannalta kaikkein tärkein kokonaisuuden kannalta. Kehitettyjä prototyyppejä pyrittiin evaluoimaan nimenomaan tästä näkökulmasta.

Munznerin [2009] mukaan evaluoinnit voidaan jakaa tavoitteidensa perusteella kolmelle eri tasolle. **Muodollisten** (formative) evaluointien tarkoitus on tuottaa suuntaviivoja suunnittelijoille visualisoinnin kehittämiseen entistä paremmaksi ja käytettävämmäksi. **Summatiivisessa** (summative) evaluoinnissa keskitytään järjestelmän suorituskyvyn ja ominaisuuksien oikeellisuuden arvioimiseen useampien testihenkilöiden avulla. **Tutkivassa** (exploratory) evaluoinnissa pyritään syventämään ymmärrystä visualisoitavasta datajoukosta ja kohdealueesta. Muodollisissa evaluoinneissa pyritään vastaamaan kysymykseen, voiko visualisointia parantaa, summatiivisessa pohditaan, onko visualisointi tehty oikein ja tutkivassa, voinko ymmärtää visualisoinnista vielä enemmän. Andrews [2008] esittelee näiden kolmen luokan lisäksi vielä ennustavan (predictive) evaluoinnin. Sen tarkoituksena on yksittäisten koehenkilöiden suoriutumisen perusteella ennakoida laajempien käyttäjäryhmien käyttäytymistä liittyen evaluoitavana olevaan visualisointiin.

Näitä evaluoinnin eri tasoja voidaan toteuttaa visualisointiprosessin yksittäisissä vaiheissa samoin kuin visualisointiprosessin lopuksi tehtävässä kokonaisevaluoinnissakin [Munzner, 2009]. Visualisointiprosessin kaikissa vaiheissa tutkiva evaluointi tukee datajoukon ja sen merkitysten ja suhteiden ymmärryksen lisääntymistä. Prototyypeille tehty laadullinen evaluointi puolestaan on luonteeltaan yhdistelmä Munznerin [2009] luokituksen muodollisesta ja summatiivisesta tasosta. Tarkoituksena oli sekä validoida tehtyjä ratkaisuja kohdealueen asiantuntijan avulla, että löytää kriittisimpiä kehityskohteita varsinaiseen toteutukseen. Evaluointiin valittiin vain pieni joukko kohdealueen asiantuntijoita, joten summatiivisen evaluoinnin edellyttämä laajuus ei tässä esimerkkivisualisointiprosessissa toteutunut.

Shneidermanin ja Plaisantin [2006] kehittämä moniulotteinen syvälinen pitkittäistapaustutkimusmenetelmä (multi-dimensional in-depth long-term case study) on uudentyypinen evaluointimenetelmä, joka on kehitetty erityisesti tiedon visualisoinnin tarpeisiin. Moniulotteisuudella tarkoitetaan käyttäjäpalautteen keräämistä monipuolisilla menetelmillä kuten havainnoinnilla, haastatteluilla ja kyselyillä mutta myös teknisillä apuvälineillä kuten automatisoiduilla käyttäjän toimintojen kirjauksella (automated logging). Syvälinisyydellä tarkoitetaan visualisoinnin suunnittelijan intensiivistä yhteistyötä kohdealueen asiantuntijan kanssa. Siinä käyttäjä ja suunnittelija pyrkivät yhteistyökumppaneina samaan päämäärään hyödyntämällä kummankin osaamista. Evaluointi on aikaulottuvuudeltaan pitkittäistutkimus, joka alkaa käytön opastuksella ja jatkuu seuraamalla visualisoinnin käyttöä pitkällä aikavälillä. Tapaustutkimuksella korostetaan evaluoinnin keskittymistä yksittäisiin käyttäjiin heidän omassa toimintaympäristössään, kun he toimivat omien työtehtäviensä parissa.

Moniulotteinen syvälinen pitkittäistapaustutkimus olisi monessa suhteessa ihanteellinen tapa evaluoida visualisointia useasta eri näkökulmasta luotettavalla ja perusteellisella tavalla. Tällaiset tutkimukset ovat kuitenkin työläitä järjestää ja vaativat paljon resursseja, sekä suurta sitoutuneisuutta niin visualisoinnin suunnittelijalta kuin käyttäjiltä. On tärkeää pohtia visualisoinnin todellista arvoa ja merkitystä kohdealueella, jotta evaluointiin kohdennetaan tarkoituksenmukainen määrä resursseja. Esimerkkiprosessissamme luotua visualisointia ei voida pitää ainakaan tässä vaiheessa vielä niin merkittävänä innovaationa, että näin laajamittaiseen ja organisoituun tutkimukseen olisi perusteltua ryhtyä.

Esimerkkiprosessissamme syntyneistä prototyypeistä ja niiden hyödyllisyydestä pyydettiin vapaamuotoista palautetta kolmelta kohdealueen asiantuntijalta. Heillä kaikilla on vahvaa osaamista kohdealueesta sekä sen datasta. Käyttäjille toimitetuissa prototyyppien versioissa ei ollut mukana lainkaan vuorovaikutteisia toimintoja. Tämä johtuu siitä, ettei kohdeorganisaatiossa ole käytettävistä sovelluksista, jolla prototyypit toteutettiin ja jolla niitä pystyisi testaamaan. Tarkoitus oli visualisoinneista poimittujen staattisten kuvien avulla esitellä visualisointien mahdollisuuksia kohdealueella sekä mahdollistaa tehtyjen prototyyppien tarkastelu ja arvioiminen. Näin menetellen käyttäjät saivat tutustua ja tarkastella visualisointeja itsenäisesti omilla työpisteillään omassa toimintaympäristössään itselleen parhaiten sopivalla hetkellä. Prototyyppien evaluoinnilla ei haluttu kuormittaa potentiaalisia käyttäjiä tarpeettoman paljon, ja tästä syystä palautetta pyrittiin keräämään varsin kevyellä ratkaisulla.

Evaluointiin osallistuneet käyttäjät pitivät visualisointia mielenkiintoisena, havainnollisena sekä ajankohtaisena tapana tarkastella toiminnasta syntynyttä dataa. Käyttäjät näkivät perustuen annettuun palautteeseen potentiaalisia hyötyjä visualisointien käyttämisessä ammattikorkeakoulun toiminnan kehittämisen tukena. He myös arvioivat useiden muiden keskeisten toimijoiden olevan kiinnostuneita visualisoinneista ammattikorkeakoulun toimintaympäristössä. Käyttäjäpalautteessa

ehdotettiin mm., että visualisointeja esiteltäisiin tulevaisuudessa esimerkiksi opinto-ohjaajille ja koulutuspäälliköille heidän kokouksissaan.

Koska palautetta prototyyppien vuorovaikutteisten toimintojen onnistumisesta ei staattisten kuvien avulla pystytty keräämään, järjestettiin vapaamuotoisen palautteen keräämisen jälkeen yhden kohdealueen asiantuntijan kanssa erillinen evaluointisessio. Siinä asiantuntija pääsi testaamaan prototyyppien vuorovaikutteisia versioita suunnittelijan testiympäristössä lyhyen alkuopastuksen jälkeen. Asiantuntijaa pyydettiin ajattelemaan ääneen tarkastellessaan visualisointeja ja häntä havainnoitiin evaluointisession ajan. Hänen havaintonsa, palautteensa ja kehitysehdotuksensa sekä havainnoijan tekemät huomiot kirjattiin ylös jatkokehitystä varten. Ääneen ajattelun ja yksittäisen henkilön havainnoinnin avulla saadut tulokset eivät ole sellaisenaan yleistettävissä laajemmin. Yksittäisen visualisointiprosessin näkökulmasta suora asiantuntijapalaute on kuitenkin todella arvokasta prototyyppien jatkokehittämiselle varsinaisiksi käyttöönotettaviksi toteutuksiksi.

Ääneen ajattelu- ja havainnointisession avulla saatiin palautetta liittyen erityisesti visualisointiprosessin toisessa ja kolmannessa vaiheessa tehtyihin ratkaisuihin. Näihin vaiheisiin palattiin evaluoinnin jälkeen iteratiivisesti tekemään muutoksia ja kehitettiin ensimmäisten prototyyppien rinnalle myös toiset vaihtoehtoiset prototyypit. Evaluointi osoitti toteen van Wijkin [2006 I] väittämän siitä, että ensimmäisenä mieleen tuleva ratkaisu harvoin on ainut mahdollinen tai edes paras toteutustapa. Aikaisten prototyyppien käyttö ja niiden evaluointi kohdealueen asiantuntijoilla osoittautui onnistuneeksi ratkaisuksi. Tämä johtuu siitä, että vaihtoehtoisten ratkaisumallien luominen prototyypeille on resurssien tehokkaan käytön kannalta merkittävästi kannattavampaa kuin lähteä isoihin näkökulman muutoksiin tai toisenlaisiin toteutustapoihin vasta varsinaisen toteutuksen jälkeen.

Vastoin Tuften [1983] laajalle levinnyttä käsitystä siitä, että apuruudukot visualisoinneissa ovat vain ylimääräistä mustetta ja käyttäjää häiritsevää kohinaa, esimerkkiprojektin visualisointiprosessissa saadussa palautteessa nimenomaan kaivattiin apuruudukkoa kaavioiden taustalle helpottamaan visuaalisten elementtien esittämien arvojen tulkintaa. Tämän palautteen perusteella prototyypeihin lisättiin jälkikäteen himmeät apuruudukot, jotka alkuperäisistä prototyypeistä oli jätetty pois kirjallisuuden kehotusta noudattaen. Esimerkkimme pienellä otoksella tehdyn evaluoinnin perusteella ei kuitenkaan voida yleisemmin väittää, että kaikissa sovelluksissa apuruudukoista olisi enemmän hyötyä kuin haittaa. Esimerkkiprojektin visualisointiprosessissamme niiden käyttö kuitenkin osoittautui perustelluksi.

Vuorovaikutteisista toiminnoista annettu palaute oli positiivista. Evaluointisessiossa havaittiin koekäyttäjän suosivan erityisesti valintaoperaatiota. Sen avulla hän pystyi korostamaan yksittäisen kulloinkin tarkasteltavana olevan viivan tai palkin keskeisyyttä muiden elementtien himmentyessä taustalle viiteinformaatioksi. Tarkkojen arvojen näyttäminen käyttäjälle pienessä pop up –ikkunassa

visualisoinnissa todettiin tarpeelliseksi, kun käyttäjä halusi yksityiskohtaisempaa lisäinformaatiota tietyistä tekijöistä. Koekäyttöä havainnoidessa todettiin prototyypivisualisointien välillä liikkuminen käytettävyydeltään hieman kömpelöksi. Tästä syystä varsinaisessa toteutuksessa tulisi kiinnittää erityistä huomiota siirtymien sujuvuuteen visualisointien välillä.

Tutkielmassa toteutettua mittavampi useammille potentiaalisille käyttäjille järjestettävä evaluointi kannattaa kerätä vasta varsinaisen implementaation jälkeen. Tämä mahdollistaa sen, että kaikki evaluointiin osallistuvat käyttäjät pystyisivät testaamaan myös visualisointien vuorovaikutteisia toimintoja omassa todellisessa toimintaympäristössään. Prototyypeistä poiketen käyttöön otettavat visualisoinnit on toteutettava sellaisella työkalulla, joka on käytettävissä kohdealueen toimintaympäristössä. Visualisoinnit joko integroidaan osaksi kohdeorganisaation sisäistä intranettiä tai julkaistaan käytössä olevalle työryhmäalustalle.

8.3. Visualisoinnin evaluoinnin erityishaasteet

Visualisoinnin evaluoinnin tekee haastavaksi erityisesti se, että käyttäjien on usein tarkasteltava samaa dataa tai visualisointia useasta eri näkökulmasta usein pitkänkin ajan kuluessa, ennen kuin he löytävät siitä kaiken itselleen tai toimintaympäristölleen olennaisen. Uusia kysymyksiä ja ideoita voi herätä vasta pitkänkin ajan päästä siitä, kun visualisointia on ensimmäisen kerran tarkasteltu. Tähän haasteeseen ratkaisuksi on esitetty pitkittäistutkimuksien käyttöä evaluointimenetelmänä. [Plaisant, 2004]

Suuret uudet löydökset ovat visualisoinneissa harvinaisia, joskin sitäkin toivottavampia löytyessään. Toisin sanoen pelkästään niiden perusteella ei visualisoinnin onnistumista kannata arvioida. Visualisointien yksi tehtävä on toisinaan vastata myös sellaisiin kysymyksiin, joita käyttäjä ei ollut itse muodostanutkaan. Myös tämä tehtävä muodostaa visualisointien evaluoinnille ja arvioinnille omat haasteensa. [Plaisant, 2004]. Jos visualisointi ei tuota uutta tietämystä datasta tai siinä ei ole havaittavissa mitään selkeitä hahmoja tai säännönmukaisuuksia, on vaikea arvioida mistä se johtuu. Se voi johtua siitä, että visualisointimenetelmä tai parametrien valinta käyttötarkoitukseen on ollut väärä, tai siitä, ettei datassa yksinkertaisesti ole säännöllisiä hahmoja ja uutta visualisoinnin menetelmin havaittavissa olevaa informaatiota [van Wijk, 2006 I].

Kun visualisointia käytetään tuottamaan informaatiota esimerkiksi päätöksenteon tueksi, voi toisinaan olla vaikeaa luotettavasti jäljittää, onko jokin onnistunut päätös tai kehitysaskel johtunut nimenomaan visualisoinnin tuottamasta näkemyksestä. Olisiko samaan lopputulokseen päädytty joka tapauksessa muiden tiedonlähteiden ja aiemman osaamisen perusteella? Visualisoinnin evaluoinnin vaikeus kiteytyykin seuraavaan: Visualisoinnin tavoitteena on lisätä tietämystä, jota on kuitenkin erittäin vaikea tunnistaa tai mitata. [Plaisant, 2004].

9. Yhteenveto

Tässä tutkielmassa on tarkasteltu eri näkökulmista opintoasiainhallinnon tiedon visualisointiprosessia ja sen eri vaiheiden sisältöä, merkitystä sekä esimerkkejä eri vaiheissa käytettävistä menetelmistä. Näkökulmia visualisointiprosessiin on haettu niin rinnakkaisalojen tutkimuksesta kuin vuosien varrella kirjallisuudessa esitetyistä erilaisista visualisoinnin prosessimalleistakin. Tutkielmassa on tehty synteesi erilaisista visualisoinnin prosessimalleista ja menetelmistä. Synteesin pohjalta on kehitetty tutkielman lähestymistapa visualisointiprosessiin. Tätä lähestymistapaa on sovellettu vaihe vaiheelta käyttäen esimerkkitoimintaympäristönä ammattikorkeakoulun opintoasiainhallintoa. Käyttäjälähtöisessä kehitystyössä on hyödynnetty aikaisia prototyyppejä.

Tutkielmassa esiteltiin visualisointien kannalta tärkeimpiä ihmisen havaintokyvyn piirteitä ja niitä sovellettiin suunniteltaessa esimerkkiprosessimalleja. Erilaisia visualisoinnin prosessimalleja tarkastelemalla ja arvioimalla pyrittiin lisäämään prosessiosaamista. Niiden pohjalta muodostettiin tutkielmassa käytettävä visualisointiprosessilähestymistapa käytäntöön sovellettavaksi.

Haasteina työssä on ollut tiedon visualisoinnin tutkimuksen monialaisuus, sekä kirjallisuudessa käytettävien käsitteiden epämääräisyys ja osittainen päällekkäisyys. Omat haasteensa tutkielman tekemiselle toivat myös visualisoinnin alalta yhä puuttuvat yhteisesti hyväksytyt kiistattomat tieteelliset teoriat ja mallit, sekä suhteellisen vähäinen määrä koko prosessin näkökulmasta tehtyjä tutkimuksia. Toisaalta visualisointiprosessin yksittäisistä osista löytyy todella paljon kirjallisuutta eri tutkimusalueilta. Tällöin tarkoituksenmukaisten lähteiden löytäminen ja valitseminen tiettyyn vaiheeseen osoittautui vaikeaksi. Haasteita työlle asetti myös valittu prosessinäkökulma, joka edellytti yksityiskohtaisuudeltaan saman tasoista käsittelytapaa visualisointiprosessin eri vaiheissa.

Työn tuloksena syntyi konkreettisia visualisointien prototyyppejä. Niiden pohjalta on mahdollista tulevaisuudessa toteuttaa kohdealueen asiantuntijoille ja päättäjille myös todelliseen käyttöön soveltuvia visualisointeja. Prototyyppien avulla onnistuttiin kuitenkin välittämään tietoa visualisoinnin käyttömahdollisuuksista ja potentiaalista kohdealueen asiantuntijoille. Työn edetessä syntyi selkeämpi käsitys siitä, millaisilla eri tavoilla ammattikorkeakoulun dataa pystyttäisiin visuaalisin keinoin esittämään, sekä mitä vaiheita ja toimenpiteitä onnistuneeseen lopputulokseen pääseminen käytännössä edellyttää.

Tutkielman ohessa syntyi myös lukuisia uusia aihioita, joita kehittelemällä voitaisiin toteuttaa visualisointeja myös kohdealueen muihin kuin tässä tutkielmassa käsiteltyihin tarpeisiin. Näin visualisointiosaamista voitaisiin kehittää konkreettisten prosessien avulla edelleen myös

tulevaisuudessa ja jalkauttaa visualisoinneista saatavia hyötyjä osaksi kohdeorganisaation jokapäiväistä toimintaa.

Viiteluettelo

- [Abela, 2009] Andrew Abela, Available: www.ExtremePresentation.com, 2009. Viitattu 6.10.2015.
- [Ackoff, 1989] Russell Ackoff, From data to wisdom. *Journal of Applied Systems Analysis*. **16** (1989), 3-9.
- [Amar et al., 2005] Robert Amar, James Eagan & John Stasko, Low-level components of analytic activity in information visualization. In: *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2005)*. (2005), 15.
- [Andrews, 2008] Keith Andrews, Evaluation comes in many guises. In: *Proceedings of the BELIV 2008 – Beyond time and errors: novel evaluation methods for Information Visualization, A Workshop of the ACM CHI 2008 Conference*. Position Paper. (2008).
- [Bertin, 1977] Jacques Bertin, *Graphics and Graphic Information Processing*. DeGruyter Press, 1977.
- [Borkin et al., 2013] Michelle Borkin, Azalea Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva & Hanspeter Pfister, What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, **18**, 1 (2013), 2306-2315.
- [Burchfield, 1972] Robert Burchfield (ed.), *Oxford English Dictionary*, Vol. **1**. Clarendon Press, 1972.
- [Card et al., 1999] Stuart Card, Jock Mackinlay & Ben Shneiderman, *Readings in Information Visualization: Using Vision to Think*. Morgan Kauffman, 1999.
- [Card, 2012] Stuart Card, Information Visualization. *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. Third edition. CRC Press, 2012.
- [Carpendale, 2008] Sheelagh Carpendale, Evaluating Information Visualizations. In: Andreas Kerren, John Stasko, Jean-Daniel Fekete & Chris North (eds.), *Information Visualization, Human-Centered Issues and Perspectives*. Springer, 2008, 19-45.
- [Chen et al., 2009] Min Chen, David Ebert, Hans Hagen, Robert Laramee, Robert van Liere, Kwan-Liu Ma, William Ribarsky, Gerek Scheuermann & Deborah Silver, Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, **29**, 1 (2009), 12-19.
- [Chittaro, 2006] Luca Chittaro, Visualizing information on mobile devices. *IEEE Computer*, **39**, 3 (2006), 40-45.
- [Christ, 1984] Richard Christ, Research for evaluating visual display codes: an emphasis on colour coding. In: Ronald Easterby & Harm Zwaga (eds.), *Information Design: The design and evaluation of signs and printed material*. John Wiley and Sons Ltd. 1984, 209-228.
- [Cleveland, 1982] Harland Cleveland, Information as Resource. *The Futurist*, (1982), 34-39.

- [Daabaj, 2000] Yousef Daabaj, The use of task analysis methods in support of the development of interactive systems. In: *Software Engineering. Proceedings of the 22nd International Conference on Software Engineering (ICSE 2000)*. (2000), 781.
- [Dix et al., 1998] Alan Dix, Janet Finlay, Gregory Abowd & Russell Beale, *Human-Computer Interaction*. Pearson Education, Prentice Hall, Harlow, 2nd edition, 1998.
- [Eliot, 1934] Thomas Stearns Eliot, *The Rock*. Faber & Faber, 1934.
- [Fayyad et al., 1996] Usama Fayyad, Gregory Piatetsky-Shapiro & Padhraic Smyth, The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM (CACM)*. **39**, 11 (1996), 27-34.
- [Fayyad et al., 2002] Usama Fayyad, Andreas Wierse & Georges Grinstein, *Information visualization in data mining and knowledge discovery*. Morgan Kaufmann, 2002.
- [firstcoastcreative, 2011] firstcoastcreative, The Five Laws. Available: <http://www.firstcoastcreative.net/gestalt-theory-graphic-design/>, 2011. Viitattu 24.8.2015.
- [García et al., 2015] Salvador García, Julián Luengo & Francisco Herrera, *Data preprocessing in data mining*. Springer, 2015.
- [Gee et al, 2005] Alexander Gee, Min Yu, Hongli Li & Georges Grinstein, *Dynamic and Interactive Dimensional Anchors for Spring-Based Visualizations*. Technical Report, Computer Science, University of Massachusetts Lowell, 2005.
- [Gershon & Eick, 1995] Nahum Gershon & Stephen Eick, Foreword. *IEEE Symposium On Information Visualization 1995 (InfoVis 1995)*. (1995).
- [González & Kobsa, 2003] Victor González & Alfred Kobsa, A workplace study of the adoption of information visualization systems. In: *Proceedings of I-KNOW'03: 3rd International Conference on Knowledge Management*. **3** (2003), 92-102.
- [Haber & McNabb, 1990] Robert Haber & David McNabb, Visualization Idioms: A Conceptual Model for Scientific Visualization Systems. In: Gregory Nielson, Bruce Shriver & Lawrence Rosenblum (eds.), *Visualization in Scientific Computing*, 1990, 74-93.
- [Hackos & Redish, 1998] JoAnn Hackos & Janice Redish, *User and Task Analysis for Interface Design*. John Wiley & Sons, 1998.
- [Han & Kamber 2006] Jiawei Han & Micheline Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2006.
- [Hansen et al., 2009] Derek Hansen, Dana Rotman, Elizabeth Bonsignore, Natasha Milic-Frayling, Eduarda Mendes Rodrigues, Marc Smith, Ben Shneiderman & Tony Capone, *Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Data*. University of Maryland Technical Report: HCIL-2009-17. 2009.
- [Harvard, 2014] Harvard School of Engineering and Applied Sciences, CS 171 Visualization, Lecture notes. Available: <https://drive.google.com/folderview?id=0BxYkKyLxfsNVN2l0Y1lFbUdlIOGc>, 2014. Viitattu 24.8.2015.

- [Hearst, 2003] Marti Hearst, *Information visualization: Principles, promise, and pragmatics*. CHI '03 Tutorial. Available: <http://www.sigchi.org/chi2003/docs/t12.pdf>, 2003. Viitattu 25.8.2015.
- [Hinterberger, 2009] Hans Hinterberger, Data Visualization. In: *Encyclopedia of Database Systems*. Ling Liu & Tamer Özsu (eds.). Springer, 2009.
- [Husa, 2015] Hannele Husa, henkilökohtainen tiedonanto, syyskuu 2015.
- [Iliinsky & Steele, 2011] Noah Iliinsky & Julie Steele, *Designing Data Visualizations*. O'Reilly, 2011.
- [Ingwersen & Järvelin, 2005] Peter Ingwersen & Kalervo Järvelin, *The Turn – Integration of Information Seeking and Retrieval in Context*. Springer, 2005.
- [Jankun-Kelly et al., 2007] T. J. Jankun-Kelly, Kwan-Liu Ma & Michael Gertz, A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, **13**, 2 (2007), 357-369.
- [Koh et al., 2011] Lian Chee Koh, Aidan Slingsby, Jason Dykse & Tin Seong Kam, Developing and applying a user-centered model for the design and implementation of information visualization tools. In: *Proceedings of 15th International Conference on Information Visualisation (IV 2011)*, (2011), 90-95.
- [Komlodi et al., 2004] Anita Komlodi, Andrew Sears & Enrique Stanziola. *Information visualization evaluation review*. ISRC Tech. Report, Dept. of Information Systems, UMBC, 2004.
- [Korkeakoulujen tietomalli] Korkeakoulujen tietomalli. Saatavilla: <http://tietomalli.csc.fi>. Viitattu 21.9.2015.
- [Kosslyn, 1985] Stephen Kosslyn, Graphics and human information processing: A review of five books. *Journal of the American Statistical Association*, **80**, 391 (1985), 499-512.
- [Larrea et al., 2007] Martin Larrea, Servio Martig & Silvia Castro, Semantic based visualization. *IX Workshop de Investigadores en Ciencias de la Computación*. (2007).
- [Lam, 2008] Heidi Lam, A framework of interaction costs in information visualization. *IEEE Transactions on Visualization and Computer Graphics*. **14**, 6 (2008), 1149-1156.
- [Mackinlay, 1986] Jock Mackinlay, Automating the design of graphical presentations of relational information. *ACM Transactions On Graphics (TOG)*. **5**, 2 (1986), 110-141.
- [Malyanov et al., 2013] Ilya Malyanov, Brian d'Auriol & Sungyoung Lee, Visualization experience and related process modeling. *Journal of Visual Languages and Computing*. **24**, 4 (2013), 223-233.
- [Martig et al., 2003] Sergio Martig, Silvia Castro, Pablo Fillottrani & Elsa Estévez, Un Modelo Unificado de Visualización. In: *Proceeding of 9° Congreso Argentino de Ciencias de la Computación*. (2003), 881-892.
- [Merriam-Webster, 2004] Merriam-Webster Inc., *Merriam-Webster's collegiate dictionary*. Merriam-Webster, 2004.

- [Munzner, 2009] Tamara Munzner, A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*. **15**, 6 (2009), 921-928.
- [Nanowerk, 2011] Nanowerk, Designing effective figures for scientific papers. Available: <http://www.nanowerk.com/spotlight/spotid=22829.php>, 2011. Viitattu 24.8.2015.
- [Niemi, 2015] Timo Niemi, henkilökohtainen tiedonanto, marraskuu 2015.
- [Nowell et al., 2002] Lucy Nowell, Robert Schulman, & Deborah Hix, Graphical encoding for information visualization: an empirical study. In: *Proceedings of 2002 IEEE Symposium on Information Visualization (INFOVIS 2002)*. (2002), 43-50.
- [Plaisant, 2004] Catherine Plaisant, The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI 2004)*. (2004), 109-116.
- [Richardson et al., 1998] Juliet Richardson, Thomas Ormerod & Andrew Shepherd, The role of task analysis in capturing requirements for interface design. *Interacting with Computers*. **9**, 4 (1998), 367-384.
- [Roth & Mattis, 1990] Steven Roth & Joe Mattis, Data characterization for intelligent graphics presentation. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI 1990)*. (1990), 193-200.
- [Rowley, 2007] Jennifer Rowley, The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*. **33**, 2 (2007), 163-180.
- [Schwabish, 2014] Jonathan Schwabish, An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives*. **28**, 1 (2014), 209-234.
- [Shneiderman, 1996] Ben Shneiderman, The eyes have it: A task by data type taxonomy for information visualizations. In: *Proceedings of the 1996 IEEE Symposium of Visual Languages*. (1996), 336-343.
- [Shneiderman, 2001] Ben Shneiderman, Supporting creativity with advanced information-abundant user interfaces. In: Rae Earnshaw, Richard Guedj, Andries Van Dam & John Vince (eds.) *Frontiers of human-centered computing, online communities and virtual environments*. Springer, 2001, 469-480.
- [Shneiderman & Plaisant, 2006] Ben Shneiderman & Catherine Plaisant, Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: *Proceedings of the 2006 AVI Workshop on BEyond time and errors: novel evaluation methods for information visualization (BELIV 2006)*. (2006), 1-7.
- [Stevens, 1946] Stanley Smith Stevens, On the theory of scales of measurement. *Science*. **103** (1946), 677-680.
- [Tan et al., 2006] Pang-Nin Tan, Michael Steinbach & Vipin Kumar, *Introduction to Data Mining*. Addison-Wesley, 2006.
- [Tory & Möller, 2002] Melanie Tory & Torsten Möller, *A Model-based Visualization Taxonomy*. School of Computing Science, Simon Fraser University, 2002.

- [Tory & Möller, 2004 I] Melanie Tory & Torsten Möller, Rethinking Visualization: A High-Level Taxonomy. In: *Proceedings of the 10th IEEE SYMPOSIUM on Information Visualization (InfoVis 2004)*. (2004), 151-158.
- [Tory & Möller, 2004 II] Melanie Tory & Torsten Möller, Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, **10**, 1 (2004), 72-84.
- [Tufté, 1983] Edward Tufté, *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- [Tweedie, 1997] Lisa Tweedie, Characterizing interactive externalizations. In: *Human Factors in Computing Systems, CHI '97 Conference Proceedings*. (1997), 375-382.
- [Upson et al., 1989] Craig Upson, Thomas Faulhaber Jr, David Kamins, David Laidlaw, David Schlegel & Jeffrey Vroom, The Application Visualization System: A Computational Environment for Scientific Visualization. *IEEE Computer Graphics and Applications*. **9**, 4 (1989), 30-42.
- [van Wijk, 2006 I] Jarke van Wijk, Views on Visualization. *IEEE Transactions on Visualization and Computer Graphics*. **12**, 4 (2006), 421-433.
- [van Wijk, 2006 II] Jarke van Wijk, Bridging the Gaps, *IEEE Computer Graphics and Applications*. **26**, 6 (2006), 6-9.
- [Voigt et al., 2012] Martin Voigt, Stefan Pietschmann & Klaus Meißner, Towards a semantics-based, end-user-centered information visualization process. In: *Proceedings of the 3rd international workshop on semantic models for adaptive interactive systems (SEMAIS)*. (2012).
- [Ware, 2008] Colin Ware, *Visual Thinking for Design*. Elsevier, 2008.
- [Ware, 2012] Colin Ware, *Information Visualization: Perception for Design*. Morgan Kaufman, 2012.
- [Yi et al., 2007] Ji Soo Yi, Youn ah Kang, John Stasko & Julie Jacko, Toward a deeper understanding of the role of interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*. **13**, 6 (2007), 1224-1231.
- [Zeleny, 1987] Milan Zeleny, Management support systems: towards integrated knowledge management. *Human Systems Management*. **7**, 1 (1987), 59-70.
- [Ziemkiewicz & Kosara, 2007] Caroline Ziemkiewicz & Robert Kosara, *Understanding information visualization in the context of visual communication*. Technical Report CVCUNCC-07-08, 2007.