



TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

JESPERI VUORINEN
DATATIETEEN MAHDOLLISUUDET LIIKETOIMINTATIEDON
HALLINNAN KONTEKSTISSA

Kandidaatintyö

Tarkastaja: Miikka Palvalin

TIIVISTELMÄ

Jesper Vuorinen: Datatieteen mahdollisuudet liiketoimintatiedon hallinnan kontekstissa

Tampereen teknillinen yliopisto

Kandidaatintyö, 25 sivua, 0 liitesivua

Marraskuu 2018

Tekniikan kandidaatin tutkinto

Tutkinto-ohjelma: Teknis-taloudellinen, TkK

Opintosuunta: Tietojohtaminen

Tarkastaja: Miikka Palvalin

Avainsanat: Datatiede, liiketoimintatiedon hallinta, BI, koneoppiminen, DIKW-hierarkkia, hiljainen tieto, jäsenneily tieto

Työssä tutkitaan datatieteen mahdollisuuksia liiketoimintatiedon hallinnassa. Tutkimuskysymyksenä on: Missä tilanteissa liiketoimintatiedon hallintaa voidaan tukea datatieteellä? Kysymystä lähestytään seuraavien apututkimuskysymysten kautta: Millä tavoin Business Intelligence (BI) ja datatiede eroavat? Mille tiedon muodoille BI:n ja datatieteen metodeja voidaan hyödyntää? Miten koneoppiminen sopii esitettyihin malleihin?

Työssä esitetään tiedon ominaisuuksien malli, jonka avulla voidaan tunnistaa tiedon muotoja ja näin ollen tilanteita, joissa datatieteen metodeja voi hyödyntää.

Työssä päädytään tulokseen, jonka mukaan datatiede voi mahdollistaa jäsennelemättömän datan ja informaation hyödyntämisen BI:n kontekstissa, mutta perinteisen liiketoimintatiedon menetöit ovat erikoistumisensa seurauksena paremmin optimoituja jäsenneilyn datan ja informaation työstämiseen. Tämän lisäksi huomataan, että koneoppiminen voi mahdollistaa jäsenneilyn tiedon ja viisauden luonnin, ja täten rakenteellisen pääoman tyyppisen aineettoman pääoman keräämisen organisaatiossa.

ALKUSANAT

Tämä kandidaatintyö on tehty osana tietojohdamisen koulutusohjelman opintoja Tampereen teknillisessä yliopistossa. Tutkimuksen aiheena on datatieteen mahdollisuudet liiketoimintatiedon hallinnan kontekstissa.

Haluan kiittää kandidaatintyön ohjaajaa, Miikka Palvalinia hyvästä ohjauksesta työn aikana ja Tarmo Vuorista työstä saadusta palautteesta ja työn oikoluvussa auttamisesta.

Tampereella 23.11.2018

Olli Jesperi Vuorinen

SISÄLLYSLUETTELO

| | | |
|-----|---|----|
| 1. | JOHDANTO | 1 |
| 2. | TUTKIMUS JA AINEISTO..... | 3 |
| 2.1 | Tutkimuksen toteutus..... | 3 |
| 2.2 | Aineiston haku..... | 4 |
| 3. | TIEDON OLEMUS | 6 |
| 3.1 | Hiljainen ja eksplisiittinen tieto..... | 7 |
| 3.2 | DIKW-hierarkia..... | 7 |
| 3.3 | Jäsennely vs. jäsentemätön data | 10 |
| 3.4 | Tiedon ominaisuuksien malli | 10 |
| 4. | TIEDON HALLINTA JA HYÖDYNTÄMINEN | 12 |
| 4.1 | Liiketoimintatiedon hallinta | 12 |
| 4.2 | Datatiede | 14 |
| 4.3 | Koneoppiminen jäsenneltynä tietona ja viisautena | 17 |
| 5. | PÄÄTELMÄT | 19 |
| 5.1 | Miten BI ja datatiede eroavat toisistaan? | 19 |
| 5.2 | Mille tiedon muodoille BI:n ja datatieteen metodeja voidaan hyödyntää? . | 19 |
| 5.3 | Miten koneoppiminen sopii esitettyihin malleihin? | 20 |
| 5.4 | Missä tilanteissa liiketoimintatiedon hallintaa voidaan tukea datatieteellä? 20 | |
| 5.5 | Tulosten arviointi..... | 21 |
| | LÄHTEET | 23 |

KUVALUETTELO

| | |
|--|----|
| <i>Kuva 1. Koneoppimishaun tulosten läpikäynnin tarkkuus, käytetyt hakusanat ja hakujen tuloksien määrä.....</i> | 5 |
| <i>Kuva 2. Tiedon ominaisuuksien malli ja sen nimetyt elementit.....</i> | 6 |
| <i>Kuva 3. Eksplisiittinen ja hiljainen kategoria osana tiedon ominaisuuksien mallia.....</i> | 7 |
| <i>Kuva 4. DIKW-hierarkia (Rowley, 2007).</i> | 8 |
| <i>Kuva 5. DIKW-hierarkia osana tiedon ominaisuuksien mallia.....</i> | 9 |
| <i>Kuva 6. Jäsennely ja jäsenitelemätön kategoria osana tiedon ominaisuuksien mallia.....</i> | 10 |
| <i>Kuva 7. Tiedon ominaisuuksiin perustuvat jaot ja kategoriat.....</i> | 11 |
| <i>Kuva 8. BI:n suhde eri elementteihin.....</i> | 14 |
| <i>Kuva 9. Datatieteen suhde eri elementteihin.....</i> | 16 |
| <i>Kuva 10. Koneoppiminen voi mahdollistaa jäsennellyn tiedon ja viisauden elementit</i> | 18 |
| <i>Kuva 11. Yhteenveto kuvista 8-10.....</i> | 21 |

1. JOHDANTO

Nykyaikaisessa yhteiskunnassa ja taloudessa kilpailu eri toimijoiden välillä on oleellisessa roolissa kehitystä ajavana voimana. Michael Porter (1985) jakoi kilpailuedun kustannus- ja differentietuun. Kummassakin tapauksessa tieto ja osaaminen ovat avainasemassa (Olsen & Dupin-Bryant, 2016; Porter 1985), koska perinteiset tuotannon tekijät ovat tulleet laajasti kaikkien saataville. Liiketoimintatiedon hallinnalla, eli business intelligencellä (BI), jonka päämääränä on luoda ja toteuttaa liike-elämän tietojen systemaattista hankintaa, analysointia ja jakamista (Langit, 2007), on perinteisesti vastattu tähän uuteen todellisuuteen, missä tieto ja osaaminen määräävät organisaatioiden menestymisen (Eugénia et al., 2018).

BI:n rinnalle on sittemmin noussut muita tieteen aloja, jotka pyrkivät vastaamaan samoihin trendeihin omista lähtökohdistaan. Tämän voidaan nähdä johtuvan pitkälti BI:ssä vastaan tulleissa rajoitteissa, jotka ovat suunnanneet perinteisen BI:n keskittymään jäsenneilyn datan ja informaation hyödyntämiseen (Baars & Kemper, 2008). Näiden rajoitteiden seurauksena jäsennelemättömän tiedon, erityisesti hiljaisen tiedon, voidaan nähdä olevan BI:n ulottumattomissa ja usein ongelmaksi nouseekin näiden tiedon muotojen hankkiminen ja muuttaminen jäsenneiltyyn muotoon. Tämän ongelman ratkaisemista varten on esitetty useita malleja, esimerkkinä Baars & Kemper (2008).

Aiemmista malleista huolimatta Abdullah & Ahmad (2015) esittävät 5th International Conference on Electrical Engineering and Informatics -konferenssissa (ICEEI) julkaistussa paperissa, ettei BI:n alalla oltu vielä tuotettu tarpeeksi tutkimusta, jonka tulosten pohjalta jäsennelemätöntä dataa voitaisiin hyödyntää kattavasti ja luotettavasti BI:n kontekstissa (Abdullah & Ahmad, 2015). Tämän tutkimuksen motivaationa toimii juurikin tästä syntyvä tarve laajentaa BI:n käytettävissä olevan tiedon muotojen määrää ja hahmottaa mahdollisuuksia muiden alojen hyödyntämiseen BI:n kontekstissa.

Tietotekniikan kehittymisen ja sen käytön yleistymisen tuomien muutosten myötä, kuten Web 2.0, sosiaalinen media ja API-talous, tiedon merkityksen ja lähteiden lisäksi ovat monipuolistuneet myös tiedon käsittelyn ja hyödyntämisen mahdollisuudet (Rose, 2016). Datatiede onkin alana noussut BI:n rinnalle tärkeään rooliin tietoa keräävänä, jäsentävänä ja hyödyntävänä kokonaisuutena.

Van der Aalst (2016, s. 10) määrittelee datatieteet poikkitieteelliseksi alaksi, jonka päämääränä on muuttaa dataa arvoksi ja jonka käyttämä data voi olla sekä jäsenneiltyä että jäsennelemätöntä. Datatieteet toimivatkin BI:ä paremmin tilanteissa, joissa

käsitellään jäsen telemätöntä tai hiljaista tietoa (Maneth & Poulouvassilis, 2017). Datatiede onkin alana kasvanut merkittävästi niin kokonsa, työpaikkojen kuin palkkojenkin osalta, koska yhä useammat yritykset ovat siirtyneet kohti data- ja tietokeskeisiä strategioita (Olsen & Dupin-Bryant, 2016).

Datatiederatkaisujen tunnistamisen vaatima osaaminen eroaa kuitenkin monella tavoin BI-osaamisesta, eivätkä kaikki datatieteen osa-alueet ole välttämättä oleellisia BI:n harjoittajille (Van der Aalst, 2016). Vastaavasti osa datatieteen ratkaisuista johtaa samoihin lopputuloksiin kuin BI:n sisäiset ratkaisut vastaaviin tilanteisiin.

Tässä kirjallisuustutkimuksessa etsitään tilanteita, joissa datatieteellä on tarjottavaa BI:n harjoittajille ja tapoja tunnistaa tällaiset tilanteet. Tutkimus pyrkii auttamaan soveltamistilanteiden tunnistamista seuraavien apututkimuskysymysten kautta:

1. Miten BI ja datatiede eroavat toisistaan?
2. Mille tiedon muodoille BI:n ja datatieteen metodeja voidaan hyödyntää?
3. Miten koneoppiminen sopii esitettyihin malleihin?

Päätutkimuskysymyksenä on: **Missä tilanteissa liiketoimintatiedon hallintaa voidaan tukea datatieteellä?**

Tutkimus rajataan datatieteiden ja BI:n soveltamistilanteiden tunnistamiseen, eikä tarkoituksena ole tutkia yksittäisiä metodeja tai niiden käyttötilanteita, joskin metodeja esitetään laajempien sovellusmahdollisuuksien esimerkkeinä. Tutkimuksessa myös datatiede ja liiketoimintatiedon hallinta nähdään toisistaan erillisinä aloina, vaikka datatieteen metodien hyödyntäminen voitaisiin käsittää myös BI:n alan sisäisenä kehityksenä (Meredith & O'Donnell, 2011). Alat erotetaan toisistaan, jotta perinteisen BI:n ja datatieteen mahdollisuuksien eroja olisi helpompi hahmottaa.

Apuna tutkittavan aiheen kohdentamisessa käytäntöön käytetään tutkimuksessa kehitettyä tiedon ominaisuuksien mallia. Malli yhdistää tiedon konseptuaalisen jaon eksplisiittiseen ja hiljaiseen tietoon, jäseneltyyn ja jäsen telemättömään tietoon sekä DIKW-hierarkiaan. Jakojen tarkoitus on helpottaa sellaisten käytännön tilanteiden tunnistamista, joissa tutkimuksessa tehtyjen havaintojen hyödyntäminen on mahdollista. Esitettyä mallia vastaavaa kuvausta ei tutkimusta tehtäessä löydetty, joskaan käytetty aineisto ei ole itsessään riittävän laaja vastaavan mallin olemassa olon mahdollisuuden poislukemiseksi.

2. TUTKIMUS JA AINEISTO

Tässä luvussa käsitellään tutkimuksen toteutustapa, aineiston haku ja tulosten vertaamista kirjallisuuteen. Aineiston haku on tehty tutkimukselle aiemmin määriteltyjen rajausten mukaan.

2.1 Tutkimuksen toteutus

Tutkimus on suoritettu kirjallisuustutkimuksena, jonka pääasiallisina hakuportaaleina toimivat Andor, Science Direct, Scopus ja Web of Science. Tutkimuksessa käytetyt lähteet on pyritty rajamaan mahdollisimman uusiin julkaisuihin, joskin vanhempiakin lähteitä on valittu mukaan erityisesti niiden ollessa termien ja aihepiirien määrittelyn kannalta oleellisia. Tämä lähdeaineistoa koskeva rajausta perustuu erityisesti datatieteiden nopeaan kehitykseen ja tutkimuksen tavoitteeseen tuottaa ajan tasalla olevia tuloksia.

Tutkimus suoritettiin osana Tampereen teknillisen yliopiston (TTY) tekniikan kandidaatin tutkintoa. Tutkimus prosessi jakautui Vaismaan (2009) esittämän mallin tavoin neljään osaan: Aiheeseen perehtymiseen, tutkimuksen suunnitteluun, tutkimussuunnitelman toteutukseen ja tutkimuksen raportointiin.

Aihepiirin valinta tapahtui tutkimuksen tekijän tiedon olemusta ja hyödyntämistä kohtaan kokeman mielenkiinnon perusteella ja aiheen rajaaminen suoritettiin keskittymällä kontekstiin, jossa tietoa ja sen olemusta voidaan hyödyntää. Aiheeseen perehdyttiin tutkimalla aiheelle olleellista kirjallisuutta aiemmin mainittujen hakuportaalien ja TTY:n tarjoamien kurssien materiaalien pohjalta. Tutkimuksen yhteydessä luotiin aiheanalyysi TTY:n kurssimuotoista kandidaatintyön suoritusta varten.

Tutkimuksen suunnittelussa luotiin alustava tutkimusaikataulu ja tutkimuksen etenemistä kuvaava rakenne. Aineiston keruuta käsitellään tarkemmin Aineiston haku - osiossa ja **kuvassa 2**. Tutkimukselle laadittiin myös tutkimussuunnitelma, aiheanalyysin tavoin osana kandidaatintyö kurssia.

Tutkimussuunnitelman toteutuksessa jouduttiin tekemään useita muokkauksia alkuperäiseen tutkimussuunnitelmaan, johtuen datatiedettä ja BI:ä yhdenaikaisesti käsittelevän tutkimuksen ja kirjallisuuden puutteesta.

Tutkimuksen raportointi suoritetaan tämän kandidaatintyön ja TTY:n kandidaatintyön kurssin sisältämien seminaarien kautta. Tämä kandidaatin työ on kuitenkin myös

itsenäinen kokonaisuus, sillä seminaarit ovat luonteeltaan vertaisarvioivia, eikä tutkimuksen tulosten ymmärtäminen vaadi tietoa näistä seminaareista.

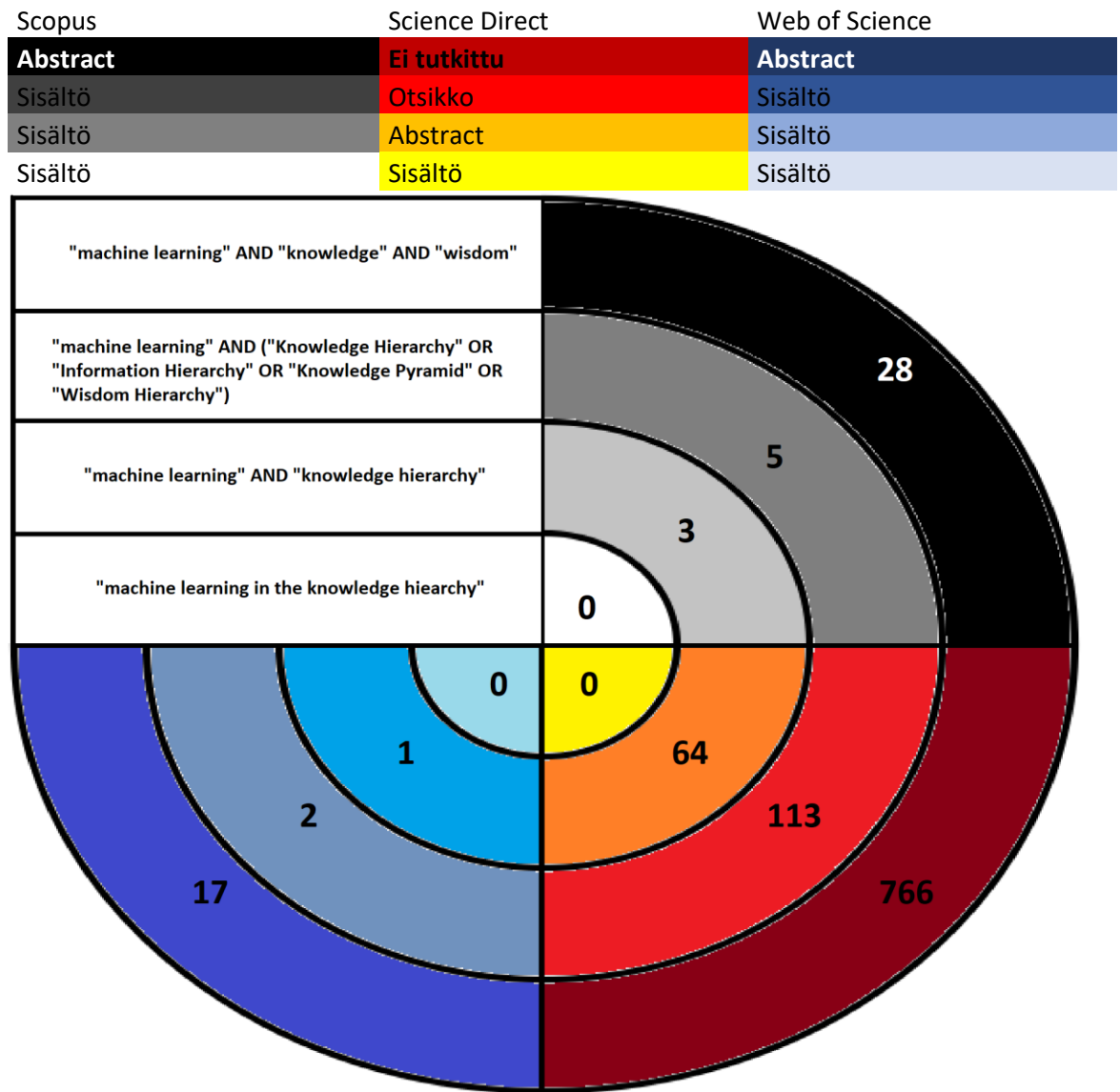
2.2 Aineiston haku

Tutkimusaineistoon perehdyttiin tekemällä järjestelmällisiä hakuja aihepiireihin liittyen. Tämän tuloksena päädyttiin päätelmään, että käsitellyt aiheet eivät ole akateemisesti riittävän läheisiä, jotta hakuja voitaisiin suorittaa muutamaaan laajempaa hakuja rajaamalla. Tämän takia päädyttiin tekemään yksittäisiä, aihekohtaisia hakuja jokaiseen käsiteltävään aiheeseen liittyen.

Jotta datatiedettä ja BI:ä olisi helppo verrata toisiinsa eri tiedon muotojen kontekstissa, käsitetään BI:n kontekstissa tapahtuva datatiedeprosessi tässä tutkimuksessa BI:n tavoin kolme vaihetta ja kolme toiminnon tasoa käsittävänä prosessina, minkä kautta myös tulkitaan datatieteitä käsitteleviä aineistoja. Käytetyt aineistot BI:n ja datatieteen osalualueilta ovat myös pitkälti toisistaan erillisiä, ja tästä johtuen joudutaan aineistoja tulkitsemaan omien lähtökohtiensa lisäksi myös toistensa kontekstissa.

Uutta tiedon ominaisuuksien mallia kehitettäessä todettiin, että koneoppiminen osoittautuu monella tavalla mallin kyseenalaisimmaksi osuudeksi. Tästä johtuen suoritettiin tutkimuksen yhteydessä laajempi aihepiirikatsaus koneoppimiseen ja sen suhteeseen DIKW-hierarkian kautta uuteen malliin tiedon ominaisuuksista. Aihepiirikatsaus suoritettiin Science Direct, Scopus ja Web of Science portaalien kautta.

Hakuportaali-kohtaiset tulosten määrät tärkeimmillä käytetyillä hakusanoilla löytyvät **kuvasta 1**. **Kuvassa 1** on myös värikoodattu tarkkuuden tasot, joilla hakutulosten sisältö on käyty läpi. Knowledge hierarchy -synonyymeinä on käytetty Rowleyn (2007) kokoamia synonyymejä. Tutkimuksen tuloksia on verrattu näissä hauissa saatuihin tuloksiin. Tuloksien perusteella tässä tutkimuksessa esitetty koneoppimisen suhde DIKW-hierarkiaan on linjassa uudemman kirjallisuuden kanssa, ja vanhemmassa kirjallisuudessa esiintyviä ristiriitaisuuksia (kuten Rowley (2007)), on myös otettu huomioon.

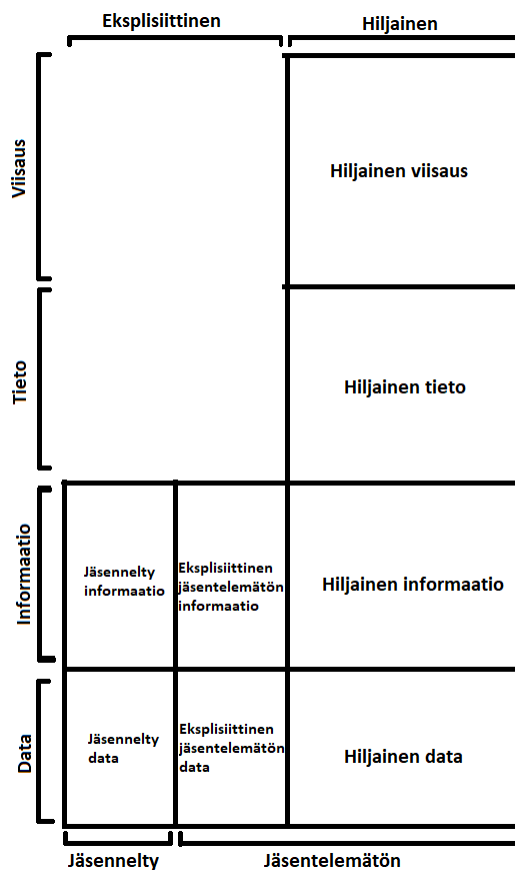


Kuva 1. Koneoppimishaun tulosten läpikäynnin tarkkuus, käytetyt hakusanat ja hakujen tuloksien määrä

3. TIEDON OLEMUS

Liiketoiminnassa tiedon ja osaamisen merkityksen kasvaessa kasvaa myös tiedon olemuksen ymmärtämisen arvo. Niin BI:ssä kuin datatieteessä käytetään useista eri lähteistä saatavaa informaatiota ja se on pystyttävä luokittelemaan olemuksensa perusteella, jotta voidaan valita tilanteeseen sopivat menetöt. Tämä on totta myös BI:n harjoittajien miettiessä: Onko datatieteellä tarjottavaa tämän tilanteen käsittelyssä?

Tässä luvussa esitetään uusi tiedon ominaisuuksien malli. Mallin tarkoituksena on nimensä mukaisesti luokitella tietoa ominaisuuksiensa kautta. Mallin avulla pyritäänkin luokittelemaan tilanteita sen perusteella, mihin mallin elementtiin kuuluvaa tietoa käsitellään.



Kuva 2. Tiedon ominaisuuksien malli ja sen nimetyt elementit

3.1 Hiljainen ja eksplisiittinen tieto

Nonaka (1994) jakoi tiedon eksplisiittiseen ja hiljaiseen tietoon artikkelissaan *A Dynamic Theory of Organizational Knowledge Creation*. Sittemmin, hiljaiselle tiedolle ei ole saatu luotua yhtä, kaikkien hyväksymää määritelmää. Rowleyn (2007) määritelmässä hiljainen tieto on yksilön omaamaa tietoa, jota ei voida helposti jakaa. Tämä määritelmä on monin tavoin sama kuin Nonakan (1994) käyttämä määritelmä. Monissa tapauksissa hiljaisen tiedon voidaankin nähdä tarkoittavan tietoa, jota ei voida tunnistaa toiminnassa tai jota ei pystytä ilmaisemaan (Toom, 2012). Toom (2012) on kuitenkin sittemmin laajentanut määritelmää niin, että organisaatioiden yhteydessä hiljaisen tiedon voidaan katsoa olevan yksi organisaation piirre, kun taas liiketoiminnan yhteydessä se nähdään kaikkien osapuolien jakamana tieto- ja tapapohjana. Laajennus on tärkeä erityisesti datatieteen kannalta, sillä sen pohjalta voidaan pitää mahdollisena, että hiljaista tietoa voidaan tuoda organisaation omaamasta hiljaisen tiedon piirteestä tai eri osapuolien jakamasta tieto- ja tapapohjasta eksplisiittiseen muotoon.

Näihin määritelmiin pohjautuen tässä tutkimuksessa hiljaisella tiedolla viitataan erityisesti vaikeasti ilmaistavaan tai toiminnan pohjalta vaikeasti tunnistettavissa olevaan tietoon (Nonaka, 1994; Rowley 2007) ja tietoon, joka on hyödynnettävissä vain organisaatioon kuuluvien yksilöiden tai liiketoiminnan osapuolten toimesta (Toom, 2012). Eksplisiittiseksi tiedoksi voidaan katsoa kaikki tieto, joka ei täytä tätä hiljaisen tiedon määritelmää, eli joka on ilmaistavissa tai joka voidaan tunnistaa toiminnasta (Nonaka, 1994; Rowley, 2007). Vastaavasti organisaatioiden kontekstissa eksplisiittinen tieto voidaan nähdä tietona, joka ei ole uniikkia tai piirremäistä tietylle organisaatiolle tai joka on tuotu muotoon, joka on organisaation tai liiketoiminnan osapuolten ulkopuolelle sijoittuvien tahojen ymmärrettävissä (Toom, 2012).

Kuvassa 3 on esitetty eksplisiittisen ja hiljaisen tiedon kategorioiden sijoittuminen uuteen tiedon ominaisuuksien malliin.



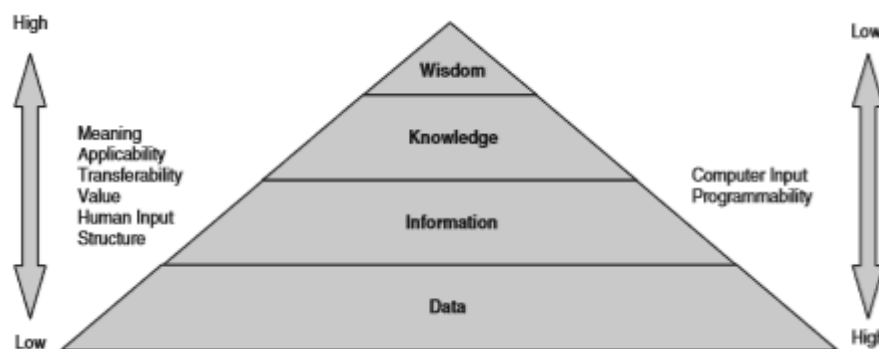
Kuva 3. Eksplisiittinen ja hiljainen kategoria osana tiedon ominaisuuksien mallia

3.2 DIKW-hierarkia

Datan, informaation ja tiedon käsitteitä käytetään arkikielessä usein toistensa kanssa keskenään vaihtokelpoisesti. Kuitenkin tietoa käsittelevän työn yhteydessä on tärkeä

tehdä ero näiden termien välille, jotta tiedon olemuksen muutosta ja niihin johtavia tekijöitä voidaan ymmärtää ja niistä voidaan käydä keskustelua (Rowley, 2007).

Journal of applied systems analysisissa julkaisemassaan artikkelissa *From data to wisdom*, Acknoff (1989) kuvailee hierarkkian, jossa data, informaatio, tieto, ymmärrys ja viisaus muodostavat rakenteen, jota usein kuvataan pyramidin muodossa. Acknoffin määritelmä on ensimmäisiä aiheen suoria kuvauksia ja sittemmin monet ovat muokanneet ja päivittäneet hierarkiaa. Erityisesti ymmärryksen asema erillisenä elementtinä on sittemmin haastettu, joten nykyään kirjallisuudessa elementeiksi muodostuvatkin useimmiten nousevassa arvojärjestyksessä vain data, informaatio, tieto ja viisaus (Rowley, 2007). Nämä neljä elementtiä ja niiden englannin kieliset termit; data, information, knowledge ja wisdom, ovat antaneet hierarkialle nimen DIKW-hierarkia, joka on esitetty **kuva** 4. Näiden neljän elementin lisäksi ehdotetut elementit eivät myöskään rajoitu samalle hierarkian tasolle Acknoffin (1989) ymmärryksen kanssa, vaan esimerkiksi Zeleny (1987) ehdotti viisauden yläpuolelle sijoitettavaa valaistumisen elementtiä.

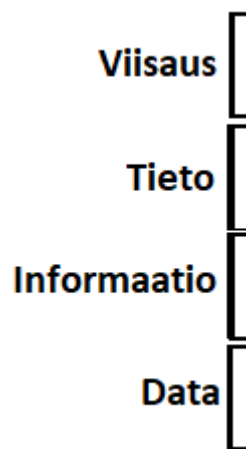


Kuva 4. DIKW-hierarkia (Rowley, 2007).

Hierarkiasta onkin esitetty useita eri malleja, mutta Rowley (2007) esittää hierarkian representaatioita kuvaavassa julkaisussaan, että yleisesti ottaen representaatiolle on neljän peruselementin lisäksi yhteistä, että ylemmät elementit voidaan esittää alempien elementtien avulla ja elementtien välillä voidaan tunnistaa muunnoksia, joiden avulla nämä ylemmät elementit selitetään. Lisäksi voidaan tunnistaa muutoksia, joiden avulla alempien elementtien sisältö, kuten data, voidaan muuttaa korkeamman elementin sisällöksi, esimerkiksi informaatioksi, tiedoksi tai viisaudeksi. (Rowley, 2007.)

Ylemmät elementit määritelläänkin usein alempien elementtien avulla (Rowley, 2007), mutta tästä huolimatta määritelmien välillä on eroja. Acknoff (1989) määrittelee datan symboleiksi, jotka edustavat ominaisuuksia, kun taas Hirschheim et al (1995)

määrittelevät dataksi väittämän, joka voi olla totta tai valheellinen. Informaatiota Acknoff (1989) kuvailee datasta kootuksi kokonaisuudeksi, jonka avulla voidaan vastata kuka, mitä, missä ja milloin tapahtuu. Hirschheim et al (1995) puolestaan kuvaavat informaation väittämänä (eli datana) jolla on lisäksi tarkoitus ja tiedon tarkoituksellisenä (informaatio) väittämänä (data), jolla on perusteltu pohja todellisuudessa. Acknoffin (1989) mukaan tiedon ominaispiirre on kyky vastata siihen, miten jokin tapahtuu ja viisaus puolestaan vastaa luotettavasti kysymyksiin siitä, miksi jotain tapahtuu. Hirschheim et al (1995) määritelmät ovat selvästi Acknoffia nykyaikaisempia ja muut aiheesta sittemmin kirjoittaneet, kuten Rowley (2007) ja Kettinger & Li (2010) ovat lähempänä Hirschheim et al (1995) määritelmiä. DIKW-hierarkia osana tiedon ominaisuuksien mallia on esitetty **kuvassa 5**.



Kuva 5. DIKW-hierarkia osana tiedon ominaisuuksien mallia

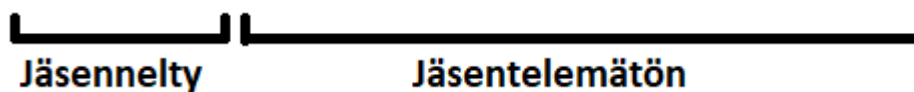
Hierarkian elementtien väliset suhteet ja olemus tarkoittavat myös, että elementtien ominaisuudet muuttuvat hierarkiassa ylös tai alas liikuttaessa (Rowley 2007). Tälle tutkimukselle oleellimmat ominaisuudet ovat elementtien arvo ja ohjelmoitavuus. Rowleyn (2007) ja Batran (2014) mukaan elementtien sisällön ohjelmoitavuus laskee hierarkiassa ylös siirryttäessä. Samanaikaisesti elementtien arvo nousee. Eri elementtien ominaisuuksien suhde niiden sijaintiin hierarkiassa on esitetty **kuvassa 4**.

On hyvä huomata, että kuten Ackoffin (1989) esittämän ymmärryselementin ja tiedon kanssa, myös informaation ja tiedon välinen ero on kyseenalaistettu. Näin erityisesti tapauksissa, joissa informaation tai tiedon tallennus alustana toimivat ihmismielten sijaan tietotekniset ratkaisut, kuten tietokannat (Batra, 2014). Tämä on oleellista erityisesti datatieteiden yhteydessä, sillä datatieteiden metodit perustuvat usein ohjelmoitavassa muodossa oleviin elementteihin.

3.3 Jäsennelty vs. jäsen telemätön data

Ohjelmoitavuuteen liittyy oleellisesti myös jäsennellyn ja jäsen telemättömän datan käsitteet. Baars & Kemper (2008), määrittivät Information Systems Managementissa julkaisemassaan artikkelissa jäsennellyn datan olevan dataa, joka on esitetty muodossa, jota pystytään suoraan prosessoimaan tietotekniikan menetelmin. Jäsen telemätön data onkin tämän vastakohta, eli dataa, jota ei pystytä sellaisenaan hyödyntämään ja prosessoimaan (Baars & Kemper, 2008). Kuten DIKW-hierarkian elementtien kanssa, myös jäsen telemätöntä dataa voidaan muuttaa muotoon, joka täyttää jäsennellyn datan vaatimukset (Baars & Kemper, 2008).

Toomin (2012) eksplisiittisen ja hiljaisen tiedon mukaan jäsennelty tieto onkin Baars & Kemperin (2008) mukaan aina myös eksplisiittistä. Näin ollen jäsennelty tieto voidaankin nähdä eksplisiittistä tietoa tarkemmin rajattuna kategoriana, jossa vaatimuksena on myös kyky käsitellä tietoa ohjelmallisesti. Rowley (2007) tuli myös lopputulokseen, jossa eksplisiittistä tietoa ei voida nähdä olevan mitään muuta, kuin informaatiota, jos tieto ja viisus nähdään ihmisten mieliin ja ajatusmalleihin rajoittuneena elementtinä. Jäsennelty ja jäsen telemätön kategoriat osana tiedon ominaisuuksien mallia on esitetty **kuvas**a 6.



Kuva 6. Jäsennelty ja jäsen telemätön kategoriat osana tiedon ominaisuuksien mallia

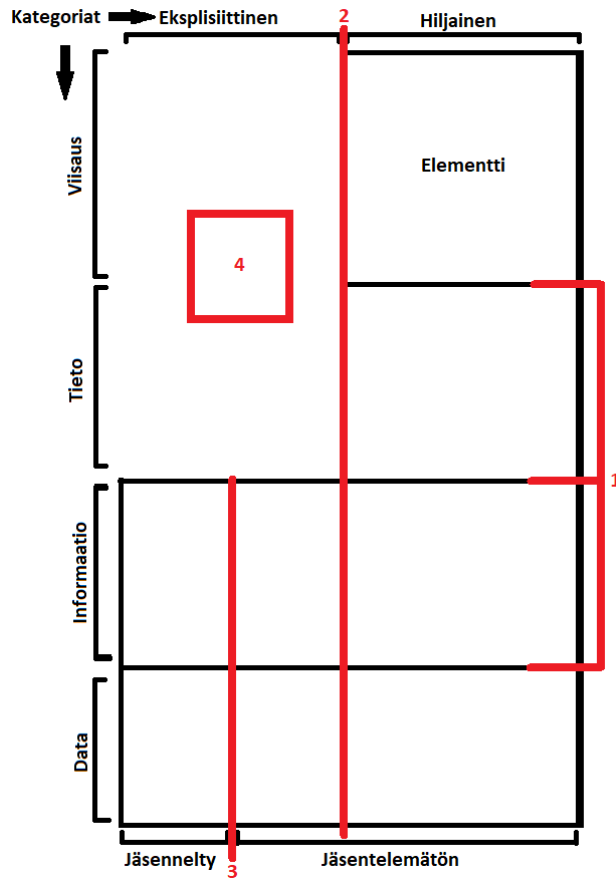
3.4 Tiedon ominaisuuksien malli

Tiedon eksplisiittisyys ja hiljaisuus, jako DIKW-hierarkian elementteihin ja jäsennelty-jäsen telemätön ulottuvuus ovatkin monin tavoin liitoksissa toisiinsa. Näiden ominaisuuksien kuvaamista varten voidaan luoda malli, joka perustuu neljään väittämään:

1. Data, informaatio, tieto ja viisaus voidaan nähdä kategorioina samassa hierarkkisessa rakenteessa (Acknoff, 1989; Hirschheim et al 1995; Zeleny, 1987).
2. Kategoriat voidaan jakaa kahteen alakategoriaan: Hiljaiseen ja eksplisiittiseen kategoriaan Nonaka, 1994; Rowleyn 2007).
3. Eksplisiittinen kategoria voidaan jakaa jäsenneltyihin ja jäsen telemättömiin elementteihin. Hiljaisen kategorian elementit ovat luonteeltaan jäsen telemättömiä (Baars & Kemper 2008; Toom, 2012).

4. Tieto ja viisaus ovat ihmismielille ominaisia kategorioita, eikä eksplisiittisen tiedon ja viisauden elementtejä voi täten olla olemassa (Rowley 2007).

Näin luotu malli on tiedon ominaisuuksien malli. Mallissa on nähtävissä tiedon jako kategorioihin ja elementteihin näiden väittämien perusteella ja väittämien luomat jaot. Malli jakoineen ja kategorioineen on esitetty **kuvassa 7** ja elementit on nimetty **kuvassa 2**.



Kuva 7. Tiedon ominaisuuksiin perustuvat jaot ja kategoriat

4. TIEDON HALLINTA JA HYÖDYNTÄMINEN

Liiketoimintatiedon hallinta ja datatiede eivät ole tarkoitusperiensä ja mahdollisuuksiensa yhtenevyyksistä huolimatta toisiaan poissulkevia tai vaihtoehtoisia aloja. Usein datatiede onkin muodostunut osaksi BI:n käytössä olevia metodeita, mikä on nähtävissä myös *BI 2.0* nimityksessä, joka on ottanut vaikutteita ja pyrkinyt hyödyntämään tunnettua *Web 2.0* nimitystä (Meredith & O'Donnell, 2011). Tässä tutkimuksessa BI ja datatiede nähdään kuitenkin toisistaan erillisinä kokonaisuuksina, jotta erot perinteisen BI:n ja datatieteen välillä olisi helpompi hahmottaa.

4.1 Liiketoimintatiedon hallinta

BI:n määritelmät vaihtelevat aineistojen välillä. Journal of Information Systems & Operations Managementissa julkaistussa artikkelissa Pop & Cristina-Loredana (2013) määrittelevät BI:n ”organisaation kykynä kerätä, hallita ja organisoida dataa”, kun taas Langitin (2007) määritelmässä BI:n tarkoituksena on luoda ja toteuttaa liike-elämän tietojen systemaattista hankintaa, analysointia ja jakamista.

BI voidaan siis nähdä prosessina, jonka avulla dataa hyödynnetään ja jonka läpi kulkiessaan data jalostuu informaatioksi ja informaatiota voidaan puolestaan käyttää prosessin kautta tiedon ja viisauden luonnissa. BI-prosessin nähdään jakautuvan erillisiin vaiheisiin, joiden sisäiset toiminnot eroavat toisistaan tarkoituksensa perusteella. Vaikka BI:n määritelmässä on eroja, niin sekä Langitin (2007), että Pop & Cristina-Loredanan (2013) määritelmistä on havaittavissa kolme tarkoitukseltaan erillistä toimintoryhmää, joille tässä tutkimuksessa käytetään nimiä:

1. **Hankkivat toiminnot**
2. **Jalostavat toiminnot**
3. **Hallinnoivat toiminnot**

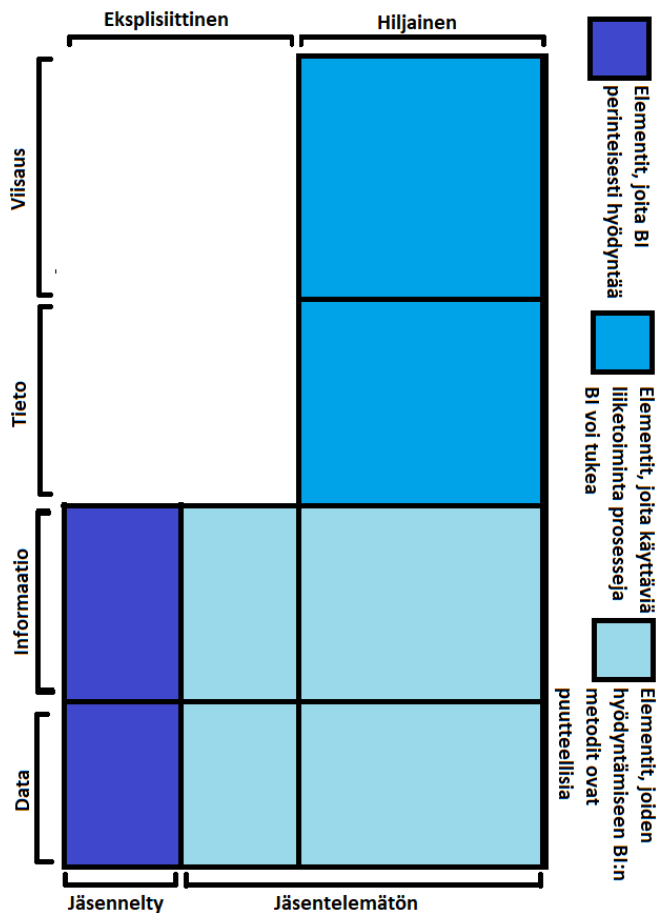
Hankkivat toiminnot ovat prosessin ensimmäisen vaiheen toimintoja. Ne tuottavat, keräävät ja hakevat dataa ja informaatiota myöhempien vaiheiden käyttöön, eivätkä itse jalosta dataa eteenpäin informaatioksi. (Langit, 2007). BI:ssä hankkivien toimintojen rajoitteena on usein niiden riippuvuus jäsenelystä datasta (Abdullah & Ahmad, 2015). Hankkivat toiminnot rajoittuvatkin toimimaan jäsenellyn datan ja jäsenellyn informaation kanssa. Täten jäsentelemätön informaatio ja data, niin hiljainen kuin eksplisiittinen, jää BI-prosessin ulkopuolelle. (Abdullah & Ahmad, 2015; Langit, 2007.)

Jalostavat toiminnot käsittelevät hankkivilta toiminnoilta saatua dataa informaatioksi, ja sijoittuvat prosessin toiseen osaan. Jalostavat toiminnot voidaan tulkita olevan Langitin (2007) määrittelemää analysointia ja Pop & Cristina-Loredanan (2013) hallintaa ja organisointia. Hankkivien toimintojen tavoin jalostavat toiminnot ovat rajoittuneet jäsenneilyn datan ja informaation kanssa toimimiseen, mutta niiden lopputuloksena voidaan jäsenneily data muuttaa jäsenneilyksi informaatioksi.

Hallinnoivat toiminnot käsittelevät, varastoivat, jakavat ja muokkaavat jalostavien toimintojen tuottamaa informaatiota prosessin kolmannessa osassa. Määritelmässä hallinnoivat toiminnot näkyvät esimerkiksi hallintana ja jakamisena (Hirschheim et al, 1995; Langit, 2007; Pop & Cristina-Loredana 2013).

Hallinnoivat toiminnot poikkeavat muista BI:n toiminnoista siinä, että niiden yhteydessä on käytössä myös hiljaista tietoa ja viisautta. Koska tieto ja viisaus ovat kuitenkin ominaisia vain ihmismielille, ovat BI-prosessin hallinnoivat toiminnot, kuten informaation jakaminen ja päätösten tukeminen, vain välillisesti tekemisessä tiedon ja viisauden kanssa. (Langit, 2007; Pop & Cristina-Loredana, 2013).

Kuvassa 8 on esitetty BI-prosessin suhteet tiedon eri kategorioihin ja elementteihin. Kokonaisuutena BI onkin pitkälti kykenemätön hyödyntämään jäsenneilytöntä dataa ja informaatiota. Hiljaista tietoa ja viisautta voidaan tukea, mutta ei suoranaisesti hyödyntää ilman ihmisiin tallennetun tiedon ja osaamisen käyttöä.



Kuva 8. BI:n suhde eri elementteihin

4.2 Datatiede

BI:stä poiketen, datatieteellä ei ole pitkää historiaa akateemisena alana ja liike-elämän osa-alueena, vaan ala on vielä niin varhaisessa vaiheessa, ettei alan tarkasta määritelmästä ole vielä saatu luotua yhteisymmärrystä (Rose, 2016). Datatiede nähdäänkin usein pikemminkin joukoksi toisiinsa liittyviä metodeja ja käytäntöjä, joita voidaan hyödyntää poikkitieteellisesti (Maneth & Poulouvassilis, 2017; Rose, 2016; Van der Aalst, 2016).

BI:n kontekstissa datatiede voidaan kuitenkin nähdä BI:ä vastaavana prosessina. Tämä yhtenevyys näkyy erityisesti Van der Aalstin (2016) määritelmästä, jossa Langitin (2007) ja Pop & Cristina-Loredanan (2013) BI:n määritelmien tavoin on nähtävissä datan/tiedon hallinta, kerääminen ja muuttaminen arvoksi.

Hankkivat datatieteen toiminnot ovat BI:n hankkivien toimintojen tavoin prosessin ensimmäisen vaiheen toimintoja, BI:n hankkivien toimintojen tavoin ne tuottavat, keräävät ja hakevat dataa ja informaatiota myöhempien vaiheiden käyttöön, eivätkä itse

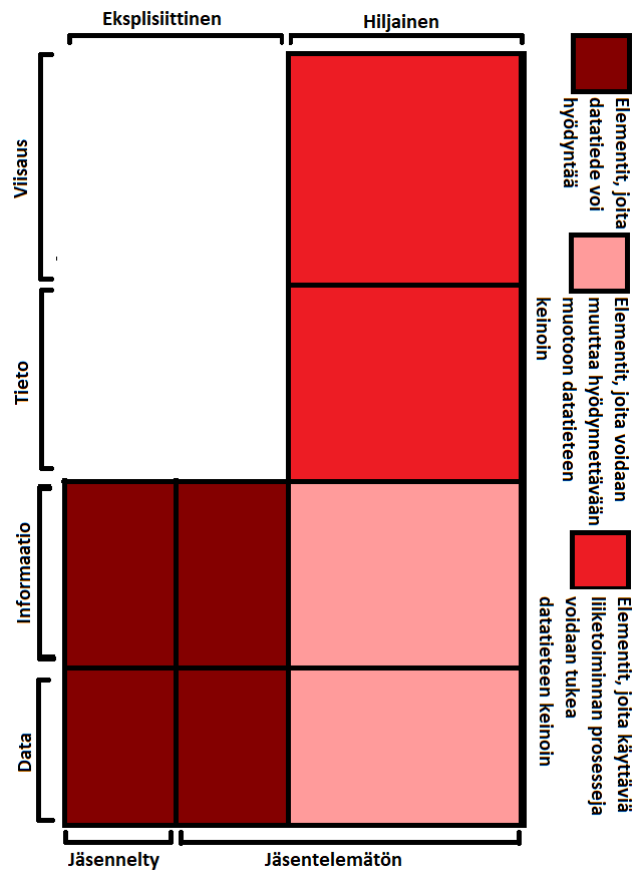
jalosta dataa eteenpäin informaatioksi (Maneth & Poulouvasilis 2017). BI:stä poiketen, datatieteen hankkivat toiminnot voivat kuitenkin hyödyntää myös jäsen telemätöntä eksplisiittistä dataa (Van der Aalst, 2016). Data cleansing tai data cleaning eli datan puhdistus on yksi datatieteen tärkeimmistä käytännöistä (Müller & Freytag, 2003; Olsen & Dupin-Bryant, 2016). Tämä puhdistus onkin oleellisessa roolissa jäsen telemättömän eksplisiittisen datan tuomisessa datatieteen hyödynnettävien elementtien piiriin. Hankkivat toiminnot eivät kuitenkaan kykene tuomaan hiljaista dataa ja informaatiota datatieteen käyttöön samassa määrin, kuin jäsen telemättömän eksplisiittisen datan ja informaation kohdalla.

Hiljaisen datan ja informaation hyödyntämisen kohdalla on tarpeen huomioida Toomin (2012) laajennettu määritelmä, jossa organisaatioiden ja liiketoiminnan osapuolten voidaan nähdä omaavan hiljaista tietoa (kuten myös dataa ja informaatiota). Organisaation kohdalla tämä hiljainen data ja informaatio on organisaation omaama piirre ja liiketoiminnassa se on osapuolten omaama tieto- ja tapapohja (Toom, 2012). Tämä data ja informaatio voidaan tuoda eksplisiittiseen muotoon informaatiotieteen ja datatieteen metodein, erityisesti Web 2.0 teknologioiden, muun muassa wikien, folksonomioiden luonnin ja taggäyksen kautta (Wagner & Bolloju, 2005; Wagner, 2004).

Wikit, folksonomiat ja taggäys toimivat periaatteella, jossa ihmisten omaamaa hiljaista tietoa tuodaan eksplisiittiseen muotoon hyödyntämällä useamman henkilön omaamaa kollektiivista tietoa. Tämä voidaan toteuttaa esimerkiksi pyytämällä tietyn organisaation jäseniä määrittelemään jokin termi tai toiminnalle oleellinen tehtävä oman omaamansa hiljaisen tiedon pohjalta, ja sitten vertailemalla kaikkien näin luotujen määritelmien jakamia yhteisiä ominaisuuksia, joista pyritään tuottamaan eksplisiittistä konensusta. (Derntl et al, 2011; Guo et al, 2009; Massimiliano, 2011; Wagner & Bolloju, 2005; Wagner, 2004) Rajatumman hiljaisen tiedon määritelmän, kuten Rowleyn (2007) ja Nonakan (1994) mukaan, tällainen hiljaisen tiedon eksplisiittiseksi muuttaminen ei onnistu näillä metodeilla.

Datatieteen jalostavat ja hallinnoivat toiminnot ovat monin tavoin yhteneviä BI:n vastaavan roolisten toimintojen kanssa (Langit, 2007; Maneth & Poulouvasilis, 2017; Pop & Cristina-Loredana 2013). Kun huomioidaan, että datatiede kokonaisuutena on BI:tä uudempi ja poikkitieteellisempi, voidaan BI:n kontekstissa toimittaessa olettaa BI:n jalostavien ja hallinnoivien toimintojen olevan erikoistumisensa seurauksena paremmin soveltuvia alalla esiintyviin tehtäviin (Van der Aalst, 2016). Jalostavat ja hallinnoivat toiminnot eivät myöskään itse hae käyttämäänsä dataa ja informaatiota, vaan saavat sen hankkivilta toiminnoilta (Pop & Cristina-Loredana, 2013). Näin ollen datatieteen jalostavat toiminnot ovat BI:n tavoin rajoittuneita jäsen nellyn datan ja informaation kanssa toimimiseen (Abdullah & Ahmad, 2015; Langit, 2007), kun taas hiljaista tietoa ja viisautta tuettaessa datatieteen hyödynnettävyys on vastaavasti rajoittunut BI:n erikoistuneisuuden ja BI-kontekstin kautta.

Kuvassa 9 on esitetty datatiede-proessin suhteet tiedon eri kategorioihin ja elementteihin. Kokonaisuutena datatiede on kykenevä tukemaan samoja hiljaisen tiedon ja viisauden elementtejä kuin BI, mutta BI:stä poiketen datatiede voi hyödyntää jäsen telemätöntä eksplisiittistä dataa ja informaatiota ja muuttaa hiljaista dataa ja informaatiota eksplisiittiseen muotoon.



Kuva 9. Datatieteen suhde eri elementteihin

Hiljaisen tiedon ja viisauden tukemisen lisäksi datatieteellä on kuitenkin yksi merkittävä uusi mahdollistava metodi tiedon ja viisauden kategorioiden kanssa työskenneltäessä; Koneoppiminen voi haastaa tiedon ja viisauden rajauksen vain ihmisille ominaisiksi kategorioiksi. Tätä mahdollisuutta tarkastellaan tarkemmin seuraavassa kappaleessa. On myös hyvä muistaa aiemmin tehty havainto DIKW-hierarkiassa tapahtuvasta arvokasvusta, kun hierarkiassa nousee ylöspäin.

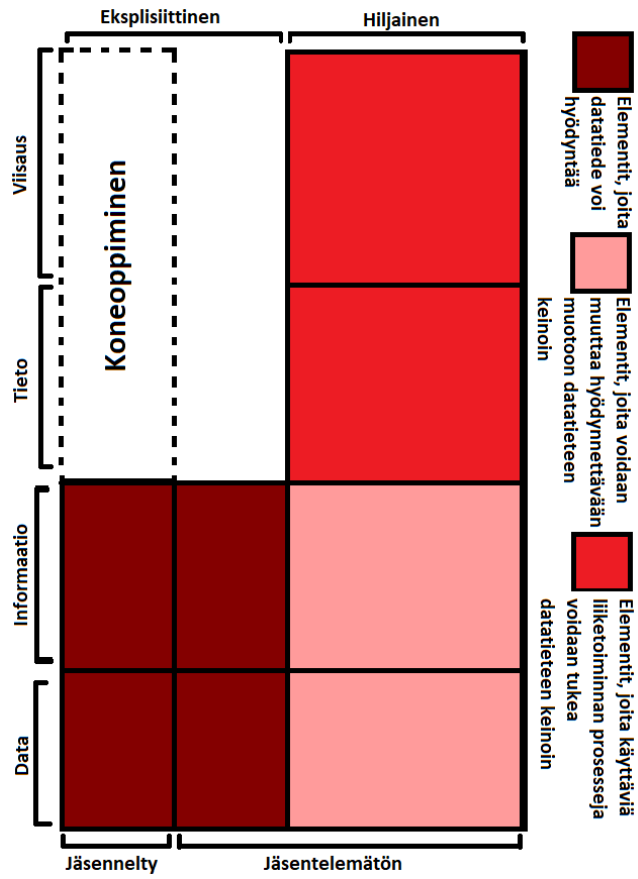
4.3 Koneoppiminen jäsennehtynä tietona ja viisautena

Koneoppiminen ei ole uusi ala, ensimmäiset koneoppimisessa käytetyt algoritmit on esitetty jo lähes 50-vuotta sitten, mutta koneoppimisen käyttötarkoitukset ja merkitys eri aloilla on kasvanut merkittävästi viime vuosina (Louridas & Ebert, 2016). Tämä muutos on johtunut tietotekniikan kehityksestä, jonka seurauksena koneoppimista on voitu hyödyntää monimutkaisissa ongelmissa, kuten data-analytiikassa ja big datan hallinnoinnissa, joissa datalla voi olla monta ulottuvuutta, mikä puolestaan luo kasvaneita vaatimuksia tietotekniikan laskentakyvyille. (Louridas & Ebert, 2016; Blei & Smyth, 2017.)

Koneoppimisen pohjana toimivat usein datasetit, joiden perusteella tietokone voi oppia hahmottamaan syy-seuraussuhteita tai suorittamaan halutun tehtävän. (Louridas & Ebert, 2016). Koneoppiminen onkin oleellinen osa datatiedettä, ja sen kyky hyödyntää dataa ja informaatiota datasettien muodossa antaa sille monia mahdollisuuksia myös BI:n alalla, missä sitä käytetäänkin jo esimerkiksi markkinoinnin ja myynnin alueilla (Louridas & Ebert, 2016; Blei & Smyth, 2017).

Koneoppiminen ei kuitenkaan sovi sellaisenaan aiempiin datatieteen toimintoihin, sillä sen toiminta täyttää selvästi niin tiedon kuin viisaudenkin määritelmät (Acknoff, 1989; Hirschheim et al 1995; Louridas & Ebert, 2016). Koneoppimisen erikoispiirteenä onkin kyky luoda ja hyödyntää tietoa ja viisautta, jota se luo jäsennehtyn informaation ja datan pohjalta. Tämä luo kuitenkin ongelman aiemmin käytettyyn tiedon ominaisuuksien malliin. Koneoppimisen on olemukseltaan ohjelmallista, tietokoneisiin ja informaatio systeemeihin tallentunutta viisautta ja tietoa, eli jäsennehtyä tietoa ja viisautta (Louridas & Ebert, 2016), jota Rowley (2007) mukaan ei tulisi olla olemassa. Koneoppiminen haastaakin käsityksemme tiedosta ja viisaudesta vain ihmisille ominaisina kategorioina.

Kuvassa 10 on kuvattu koneoppimisen luoma jäsennehty tieto ja viisaus tiedon ominaisuuksien mallissa. Koneoppimisen ja DIKW-hierarkian suhteesta tehdyn kirjallisuuskatsauksen (**Kuva 1**) perusteella näyttää siltä, että viimeaikainen kirjallisuus, kuten Ajwani et al. (2018) ja Medford A. et al. (2018) ovat poikenneet Rowleyn (2007) tapaisesta tiedon ja viisauden rajaamisesta vain ihmisille ominaisina kategorioina.



Kuva 10. Koneoppiminen voi mahdollistaa jäsennellyn tiedon ja viisauden elementit

5. PÄÄTELMÄT

Tutkimuksessa päätutkimuskysymyksenä oli: **Missä tilanteissa liiketoimintatiedon hallintaa voidaan tukea datatieteellä?** Tähän pyrittiin vastaamaan luomalla tiedon ominaisuuksia kuvaava malli (**Kuva 7**) ja hyödyntämällä tätä mallia seuraaviin apututkimuskysymyksiin vastaamisessa:

1. Miten BI ja datatiede eroavat toisistaan?
2. Mille tiedon muodoille BI:n ja datatieteen metodeja voidaan hyödyntää?
3. Miten koneoppiminen sopii esitettyihin malleihin?

5.1 Miten BI ja datatiede eroavat toisistaan?

BI ja datatiede päätettiin tutkimuksessa käsitellä toisistaan erillisinä aloina, jotta datatieteen mahdollisuuksia BI:n kontekstissa olisi helpompi hahmottaa. Tämän seurauksena BI on vaihtelevista määritelmistään huolimatta alana yhtenäinen kokonaisuus, jolle voidaan määrittää selviä päämääriä osana organisaatioiden toimintaa. BI voidaan täten kirjallisuudessa esiteltyjen määritelmien pohjalta esittää prosessina, jonka kautta kyseisiin päämääriin voidaan pyrkiä. (Abdullah & Ahmad, 2015; Langit, 2007; Meredith & O'Donnell, 2011.)

Datatiede on puolestaan monimuotoinen metodien ja ratkaisujen kokoelma, jota voidaan hyödyntää useilla eri aloilla (Maneth & Poulouvassilis, 2017; Rose, 2016; Van der Aalst, 2016). Näin ollen ala on myös BI:stä poiketen vailla yhtenäistä päämäärää ja sen tarkoitus määräytyykin käytön kontekstin kautta. Tutkimuksessa tämä konteksti oli BI-prosessi ja täten datatieteen ja BI:n päämäärät voitiinkin tuoda yhteneviksi ja määrittää datatieteelle BI:ä vastaava prosessikuvaus.

5.2 Mille tiedon muodoille BI:n ja datatieteen metodeja voidaan hyödyntää?

BI:n kontekstissa hyödynnettynä datatieteen voidaan nähdä tuovan jäsentelemättömän eksplisiittisen datan ja informaation suoraan BI-prosessin hyödynnettäväksi (Van der Aalst, 2016). Samalla hiljaisen datan ja informaation voidaan nähdä olevan epäsuorasti ja tilannekohtaisesti datatieteen hyödynnettävissä, kun apuna käytetään esimerkiksi Web 2.0 teknologioita, kuten taggäystä, wikejä ja folksonomioita, joilla organisaationaalista hiljaista tietoa voidaan tuoda jäsentelemättömän eksplisiittisen datan ja informaation muotoon, jota voidaan edelleen hyödyntää datatieteen hankkivien metodein (Wagner & Bolloju, 2005; Wagner, 2004).

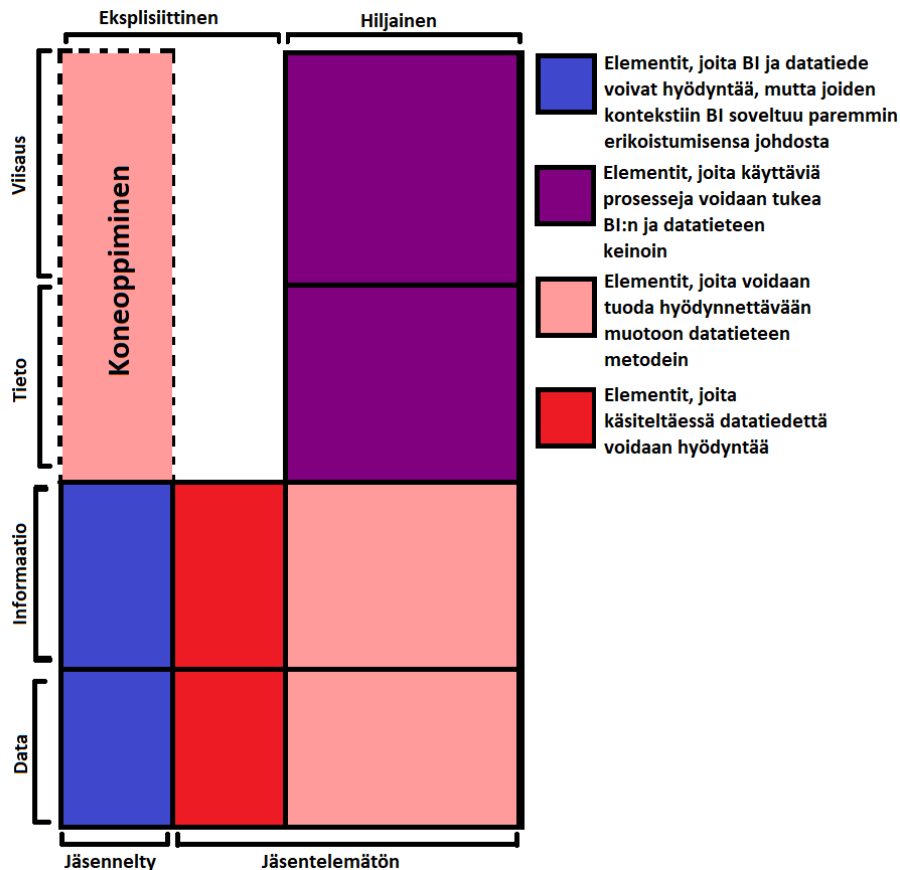
BI:llä on omat etunsa oman erikoistumisensa johdosta jäsennellyn datan ja informaation kanssa toimittaessa, hiljaista tietoa ja viisautta tuettaessa sekä jalostavia ja hallinnoivia toimintoja hyödynnettäessä. Näissäkin tilanteissa datatieteen metodit voivat olla toimivia, mutta on vaikea kuvitella tilannetta, jossa datatieteen jalostavat ja hallinnoivat toiminnot ovat BI:n vastaavan vaiheen toimintoja tehokkaampia BI:n kontekstissa toimittaessa. Todellisuudessa optimaaliset ratkaisut löytyvät mitä luultavimmin yhdistämällä BI:n ja datatieteen jalostavia ja hallinnoivia toimintoja. BI:n ja datatieteen hyödynnettävissä olevat tiedon muodot on esitetty **kuviissa 8 ja 9**.

5.3 Miten koneoppiminen sopii esitettyihin malleihin?

Siinä missä BI vaatii valmiit, ihmisen suunnittelemat rakenteet toimintansa perustaksi, pystyy datatieteeseen pohjautuva ratkaisu mukautumaan saamaansa tietoon ja etsimään itsenäisesti rakenteita kyseisestä tiedosta, esimerkiksi koneoppimisen avulla. (Louridas & Ebert, 2016; Blei & Smyth, 2017). Koneoppimisen voidaankin nähdä luovan jäsennellyä tietoa ja viisautta, johon perinteinen BI ei ole metodeineen kykenevä. Tämä on erityisen tärkeää, sillä tällainen aineettoman pääoman alle kuuluva rakenteellinen pääoma ei ole riippuvainen yksilöistä, vaan se on pysyvä osa organisaatiota. Rakenteellinen pääoma, ja siten koneoppiminen, voikin auttaa organisaatiota pysyvän, henkilöihin sitoutumattoman arvon luonnin mahdollistamisessa. Koneoppiminen osana tiedonominisuuksien mallia ja datatiedettä on esitetty **kuviissa 10**.

5.4 Missä tilanteissa liiketoimintatiedon hallintaa voidaan tukea datatieteellä?

Kuvassa 11 on esitetty yhteenveto **kuviista 8 - 10**. Kuva toimii suuntaa antavana mallina BI:n ja datatieteen toimivuuksien eroille käsiteltäessä eri elementtejä. Näiden elementtien pohjalta voidaan myös tunnistaa mihin tilanteisiin datatieteiden ja BI:n metodien hyödyntäminen parhaiten soveltuu tunnistettaessa tilanteille oleellisen tiedon ominaisuuksia.



Kuva 11. Yhteenveto kuvista 8-10.

Datatieteellä onkin paljon tarjottavaa BI:n harjoittajille. Koneoppiminen mahdollistaa uudenlaisen rakenteellisen pääoman keräämisen jäsennellyn tiedon ja viisauden luonnin kautta (**kuva 11**, vasen ylänurkka). Datatieteen hankkivat toiminnot voivat puolestaan tuoda jäsenitelemätöntä hiljaista dataa ja informaatiota ensin jäsenitelemättömäksi eksplisiittiseksi ja sitten jäsennellyksi dataksi ja informaatioksi (**kuva 11**, data ja informaatio kategorioiden sisäinen liike). Koneoppimisen kyky etsiä itsenäisesti rakenteita voi mahdollistaa myös datatieteeseen pohjautuvia hiljaisen tiedon ja viisauden tukemisen metodeja (**kuva 11**, oikea ylänurkka), jotka ovat kilpailukykyisiä BI:ssä esiintyvien erikoistuneempien vastapariensa kanssa.

5.5 Tulosten arviointi

Tutkimuksen tuloksia ei voitu suoraan verrata aiempiin tutkimuksiin, sillä datatieteiden ja BI:n suhdetta ei, tutkimuksen aineiston haun yhteydessä löydettyjen aineistojen pohjalta, ole aiemmin suoranaisesti tutkittu. Tulokset ovat kuitenkin yhdenmukaisia tutkimuksessa käsiteltyjen yksittäisten aihepiirien kirjallisuuden ja aiemman tutkimuksen kanssa. Tutkimuksessa olisi voitu käydä tarkemmin läpi BI:n ja datatieteen

sisältämiä aihepiirejä ja metodeita, mutta tämän ei nähty olevan oleellista tutkimuksen päämäärien saavuttamiseksi.

Tarkemmassa metodien ja aihepiirien läpikäynnissä ja vertailussa olisi kuitenkin mahdollisuuksia hyödylliselle jatkotutkimukselle. Samoin kehitetyn tiedon ominaisuuksien mallin syvempi tarkastelu ja sovellusmahdollisuuksien tutkiminen toisi uusia tutkimusmahdollisuuksia.

LÄHTEET

Abdullah, M., Ahmad, K. (2015). Business intelligence model for unstructured data management. 5th ICEEI, s. 10

Ackoff, R. L. (1989). From data to wisdom. Journal of applied systems analysis. Vol. 16, s. 3-9.

Ajwani, D. et al. (2018). Efficient auto-generation of taxonomies for structured knowledge discovery and organization. Proceedings of the 29th ACM Conference on Hypertext and Social Media. s. 251-252.

Baars H. & Kemper H. (2008). Management Support with Structured and Unstructured Data - An Integrated Business Intelligence Framework. Information Systems Management. Vol. 25:2. s. 132-148.

Batra, S. (2014). Big data analytics and its reflections on DIKW hierarchy. Review of Knowledge Management. Vol. 4(1). s. 5-17. Haettu: <https://search-proquest-com.libproxy.tut.fi/docview/1906365337>.

Blei D. & Smyth P. (2017). Science and data science. Proceedings of the National Academy of Sciences. Vol. 114(33). s. 8689-8692.

Chaffey, D., & White, G. (2010). Business information management: Improving performance using information systems. Pearson Education.

Derntl, M. et al. (2011). Inclusive social tagging and its support in Web 2.0 services. Computers in Human Behavior. Vol. 27(4). s. 1460-1466. Haettu: <https://www.sciencedirect.com/science/article/pii/S0747563210002906?via%3Dihub>.

Eugénia P. et al. (2018). Intellectual capital and performance: Taxonomy of components and multi-dimensional analysis axes. Journal of Intellectual Capital. Vol. 19(2). s. 407-452. Haettu: <https://doi-org.libproxy.tut.fi/10.1108/JIC-11-2016-0118>.

Guo, T. et al. (2009). Codifying collaborative knowledge: Using wikipedia as a basis for automated ontology learning. Knowledge Management Research & Practice. Vol. 7(3). s. 206-217. Haettu: <http://dx.doi.org.libproxy.tut.fi/10.1057/kmrp.2009.14>.

Hirschheim, R. et al. (1995). Information Systems Development and Data Modelling. Cambridge University Press, New York.

Kettinger, W., & Li, Y. (2010). The infological equation extended: Towards conceptual clarity in the relationship between data, information and knowledge. European Journal of Information Systems. Vol. 19(4). s. 409-421. Haettu: <http://dx.doi.org.libproxy.tut.fi/10.1057/ejis.2010.25>.

Langit, L. (2007). What is Business Intelligence? Foundations of SQL Server 2005 Business Intelligence. s. 1-24.

Louridas P. & Ebert C. (2016). Machine Learning. IEEE Software. vol. 33(5). s. 110-115. Haettu:

<http://dx.doi.org.libproxy.tut.fi/10.1057/ejis.2010.25.ieee.org.libproxy.tut.fi/stamp/stamp.jsp?tp=&arnumber=7548905&isnumber=7548893>.

Maneth, S. & Poulouvassilis, A. (2017). Data Science. The Computer Journal. Vol. 60(3). s. 285–286. Haettu: <https://doi.org.libproxy.tut.fi/10.1093/comjnl/bxw073>.

Massimiliano, D. (2011). Search for Hidden Knowledge in Collective Intelligence dealing Indeterminacy Ontology of Folksonomy with Linguistic Pragmatics and Quantum Logic

Medford A. et al. (2018). Extracting Knowledge from Data through Catalysis Informatics. ACS Catalysis. Vol. 8(8). s. 7403-7429.

Meredith, R., & O'Donnell, P. (2011). A framework for understanding the role of social media in business intelligence systems. Journal of Decision Systems. Vol. 20(3). s. 263-282. Haettu: <https://search-proquest-com.libproxy.tut.fi/docview/1314906815?accountid>

Müller, H. & Freytag, J. (2003). Problems, methods, and challenges in comprehensive data cleansing.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. Organization Science. Vol. 5(1). s. 14-37. Haettu: <http://www.jstor.org/stable/2635068>.

Olsen, D. & Dupin-Bryant, P. (2016). Integrating data cleansing with popular culture: A novel SQL character data tutorial. The Review of Business Information Systems. Vol. 20(1). Haettu: <https://search-proquest-com.libproxy.tut.fi/docview/1797558218>.

Pop, D., & Cristina-Loredana, A. (2013). Using Oracle Business Intelligence To Analyze Company Results And Create Strategies. Journal of Information Systems & Operations Management.

Porter, M. (1985). Technology and competitive advantage. Journal of Business Strategy (Pre-1986), Vol. 5. s. 60. Haettu: <https://search-proquest-com.libproxy.tut.fi/docview/209880770>.

Rose, D. (2016). Data science: create teams that ask the right questions and deliver real value.

Rowley, J. (2007) The wisdom hierarchy: representations of the DIKW hierarchy, Journal of Information Science. Vol. 33(2). s. 163–180.

Toom A. (2012). Considering the Artistry and Epistemology of Tacit Knowledge and Knowing. Educational Theory. Vol. 62(6).

Vaismaa, K 2009, Aiheesta analyysiin - tukipaketti kandidaatin- ja diplomityön tutkimusprosessiin. Tampereen teknillinen yliopisto. Tiedonhallinnan ja logistiikan laitos, Tampere.

Van der Aalst, W. (2016). Process Mining: Data Science in Action. Springer.

Wagner, C. (2004). Wiki: a technology for conversational knowledge management and group collaboration Communications of the AIS. Vol. 13. s. 265-289

Wagner, C., & Bolloju, N. (2005). Supporting knowledge management in organizations with conversational technologies: Discussion forums, weblogs, and wikis. Journal of Database Management. Vol. 16(2). Haettu: <https://search-proquest-com.libproxy.tut.fi/docview/199664960>.

Zeleny, M. (1987). Management support systems: towards integrated knowledge management. Human Systems Management. Vol. 7(1). s. 59–70.