

Arho Virkki

DI, FT

Kliinisen tietopalvelun johtaja, Turun yliopistollinen keskussairaala

Tietoaltaan avulla kohti laadukasta ja kustannustehokasta lääkehoitoa

Tietoallas sopii suurten tietomäärien yhdistämiseen. Sen avulla pääsemme analysoimaan laajoja aineistoja ja voimme löytää harvinaisiakin säännönmukaisuuksia. Tekniikka on aidosti uutta, ja tekoäly auttaa ammattilaista päätöksenteossa. Tiedon avulla kansalaiset saavat tehokkaampaa ja turvallisempaa yksilöllistä lääkehoitoa.

Tietoallas on järjestelmä, jonka avulla suuri määrä raakatie-toa voidaan tiivistää ja jalostaa päätöksentekoon sopivaksi. Teknologia syntyi 2000-luvun alussa Googlen kahden konferenssijulkaisun pohjalta. Niissä esiteltiin sekä hajautettu tiedostojärjestelmä GFS ja rinnakkaislaskentaan tarkoitettu MapReduce-algoritmi. Yahoolla työskennellyt Doug Cutting alkoi rakentaa vastavia avoimen lähdekoodin komponentteja yhdessä Mike Cafarellan kanssa, mutta projektille ei ollut aluksi nimeä. Doughin lapsella oli kuitenkin leikkielefanti Hadoop, jonka nimi oli helppo lausua eikä tarkoittanut mitään.

Mikään tekniikka ei muuta maailmaa hetkessä, mutta avoimeen lähdekoodiin ja Hadoop-järjestelmään perustuvat tietoaltaat ja massadata (big data) ovat aidosti uusi paradigma. Perinteisen ja hyödyllisen, mutta raskassoutuisen rakenteisen tietovarastoinnin rinnalle on syntynyt tekniikka, jota voidaan käyttää ketterästi suurten tietomäärien yhdistämisessä. Esimerkkejä tästä ovat potilaan monitorointi ja seuranta, tietojen yhdistäminen sadoista eri lähteistä, hahmontunnistus ja merkityksellisten muuttujien automaattinen löytäminen. Massadataa voi verrata mikroskoopin tai teleskoopin keksimiseen:

pääsemme näkemään ilmiöitä, joita ei ole aikaisemmin voitu havaita tai tutkia. Koska järjestelmä on avoin, tekniikkaan perehtyneiden on lisäksi mahdollista toteuttaa siihen parannuksia.

Tietoallas mahdollistaa lääkehoitojen räätälöinnin

Teollisuudessa tietoaltaita käytetään muun muassa toiminnanohjaukseen ja asiakasprofilointiin. Nordean konsernijohtaja Casper von Koskull toteaa Talouselämän numerossa 12.11.2017 osuvasti: ”Menemme maailmaan, jossa ei ole enää asiakassegmenttejä, vaan palveluja pystytään räätälöimään yhdelle henkilölle.” Matemaattisessa mielessä kysymys on samasta kuin hoidon kohdentamisessa yksittäiselle potilaalle.

Lääkehoitojen räätälöiminen kaikki vaihtoehdot etukäteen testaamalla on mahdotonta. Tekoölyalgoritmien avulla voimme kuitenkin havainnoida passiivisesti tuhansien muuttujien tilastollista ennustevoimaa ja käyttää tätä hyödyksi uusien päätösten yhteydessä. Tämä eroaa perinteisistä tilastollisista monimuuttujamenetelmistä kahdella tavalla: tutkittava ilmiö voi olla lähes mielivaltaisen mutkikas ja p-arvojen sijaan tutkitaan suoraan hoitovaste-ennustetta. Tekoöly, kuten siitä nykyään puhutaan tilastolliseen teoriaan perustuvana, tunnistaa datasta erilaisia ”sormenjälkiä” ja käyttää niitä päätteilyyn. Suomen nykyisiä Isaacus-hankkeissa rakennettuja tietoaltaita voi jo mainiosti käyttää datan visualisointiin ja tieteelliseen tutkimukseen. Seuraava askel on päätöksenteon tukeen suunniteltujen algoritmien ottaminen mukaan hoitavan lääkärin ja tietojohdajan työpöydälle.

Tietoaltailla on toki paljon muitakin käyttökohteita kuten toiminnan oikeellisuuden tarkistaminen ja seuranta, tietojen yhdistäminen potilaan henkilökohtaiselle aikajanelle, tasa-arvoisen hoidon toteutumisen seuranta, vaikuttavuuden ennusteet, päätöksenteon tuki ja tiedolla johtaminen. Talous- ja kiinteistödatan

yhdistäminen helpottaa hinnoittelua ja logistiikkaa. Jokaisessa sairaalan sängyssä voisi esimerkiksi olla anturi, joka tyyppillisten suureiden, kuten potilaan pulssin ja sykevariaation lisäksi, kertoisi sängyn tarkan sijainnin. Yksinkertainen kysymys saisi vastauksen: Montako sänkyä on tänään käytössä ja missä ne ovat juuri nyt?

Vastuullisesti ylläpidetyssä tietoaallassa kunnioitetaan tietosuojaa

Tietosuoja viittaa tässä käytäntöihin ja lainsäädäntöön, jonka mukaan tietoa käsitellään, kun taas tietoturva on teknistä tiedon suojaamista käyttöoikeuksin ja palomuuirein, sekä alkupeiraisen tiedon käsittelemistä siten, että yksilöä ei voida tunnistaa. Tietosuoja mainitaan usein syyksi, miksi aineistoa ei haluta yhdistettävän millään perusteella, vaikka se olisi teknisesti mahdollista. Meneillään näyttää kuitenkin olevan ajattelutapojen muutos: jos voin saada parempaa hoitoa itselleni antamalla tietoni laajempaan käyttöön, niin miksi ihmeessä kieltäisin sen. Enemmän ihmisiä tuntuu turhauttavan paperisten lomakkeiden täyttö, koska ajalla ja vaivalla kirjoitettuja vastauksia lukevat vain harvat.

Vastuullisesti ylläpidetyt tietoaltaat vievät yhteiskuntaa kohti läpinäkyvyyttä ja mahdollisuuksien tasa-arvoa. On tärkeää, että tietoturva toteutetaan matemaattisen tilastotieteen keinoin siten, että datalähtöisen tutkimuksen mahdollisuuksia ei hukata. Onnistumiseen on myös erinomaiset mahdollisuudet: Anonymisoitu (eli peruuttamattomasti tunnistamattomaksi tehty) tai fabrikoitu (satunnainen, mutta teknisesti oikean muotoinen) data sopii alkuvaiheen tuotekehitykseen ja opetukseen. Peitenimillä suojattu data, eli pseudonyymi data, soveltuu hyvin tieteelliseen tutkimukseen. Kattava yksilöity henkilötieto taas takaa kansalaiselle parhaan kohdennetun hoidon sekä ammattilaiselle päätöksenteon tuen menetelmät.

Laadukas data on kaiken perusta

Vaikka tietoallas helpottaa tiedon yhdisteltävyyttä, tehtävä on edel-

leen haasteellinen. Usein tieto sijaitsee suljetussa lähdejärjestelmässä, josta sen lukeminen on vaikeaa tai lisenssisyistä erittäin kallista. Lisäksi massadatan käsittely ja tekoölyn soveltaminen vaatii tietoteknistä ja matemaattista ammattitaitoa, jota on nykyisillä työmarkkinoilla niukasti tarjolla. Lopulta kaikki palautuu kuitenkin lähdemateriaalin luotettavuuteen. Lääke voi olla määrätty ja haettu apteekista, mutta se on jäänyt ottamatta.

Parhaimmillaan tietoallas kuitenkin kerää kattavasti eri luotettavuustasojen havaintoja sitä mukaa, kun niitä elävässä elämässä syntyy. Tässä on ero tutkimukseen, joissa tutkija päättää testattavat asiat ja järjestää niille sopivan vertailuryhmän. Lisäksi, jos terveystietoaltaasta saataisiin kansallinen, se mahdollistaisi lääkkeiden vaikutuksen seuraamisen koko populaatiossa. Iso tietovaranto voisi paljastaa harvinaisten, mutta systemaattisesti ilmenevien vakavien haittavaikutusten lisäksi yllättäviä uusia sovelluskohteita. Sopsisiko ”Tietoallas Suomen” vertailuryhmäksi vaikka Ruotsi? ●

KIRJALLISUUTTA

- Jeffrey D, Sanjay G. MapReduce: Simplified data processing on large clusters. OSDI'04: Proceedings of the 6th Conference on Symposium on Operating Systems Design and Implementation. 2004, USENIX Association.
- Kaataako Apple nyt Nordean niin kuin Nokian? – Sitä pohtii Nordean johtava von Koskull. Talouselämä. Julkaistu verkossa 12.11.2017. www.talouselama.fi
- Sanjay G, ym. The Google File System. 19th ACM Symposium on Operating Systems Principles. Lake George, NY, 10/2003.

