

Yhteentoimivuusalusta tekoälypalvelujen näkökulmasta

Marko Latvanen
erityisasiantuntija, VRK

22.8.2019



Yhteentoimivuusalusta = tietomäärittelyksiä

- Yhteentoimivuusalustan ja –menetelmän selkeä ja kirjattu tarkoitus on antaa välineet **tiedon merkityksen** määrittelylle, yhtenäistämiseksi ja merkitysten eheälle liikkumiselle organisaatioiden ja järjestelmien välisessä liikenteessä ja loppukäyttäjäpalveluissa.
- Y-alustassa ei ole mitään erityisesti juuri tekoälyjärjestelmiä varten suunniteltua ominaisuutta, joten joudumme luomaan tämän näkökulman olemassa olevan tiedon ja ymmärryksen pohjalta.



Tekoäly = ?

- Hieman kärjistäen voidaan todeta, että ns. tekoälylle ja sen tavoitteille ei ole yhtä täsmällistä määritelmää, joten keskustelu tekoälypalveluista on itsessään liikkuva maali.
- Tiedämme, että ylivoimainen kapasiteetti käsitellä ja analysoida Big Dataa on olennainen osa tekoälyä ja erottaa sen perinteisestä aatekoosta.
- Wikipedia: "Any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals ... A system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation."
- Amazon: "The field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem solving, and pattern recognition"



Yhteentoimivuusalusta + AI = ?

- "perceives", "takes actions", "correctly interprets external data", "learns from such data", "flexible adaptation" = jotta AI olisi "oikea" AI, se tarvitsee erittäin paljon dataa, järeät laskentatehot, kyvyn ottaa vastaan (tai noutaa) ja prosessoida itsensä ulkopuolista dataa, jonka oppimisen myötä se sopeutuu toimintaympäristöön ja uusiin tehtäviin (autonomiset ominaisuudet).
- **Miten Yhteentoimivuusalusta liittyy tähän?** Jotta voimme sanoa, että Yhteentoimivuusalustan työkalut ja ominaisuudet tukevat, edistävät tai kiihdyttävät juuri tekoälypalvelujen kehittämistä, niiden olisi vaikutettava yllä kirjattuihin AI-järjestelmien ominaisuuksiin.



Mistä puhumme, kun puhumme tekoälypalveluista? 1/2

- Ennen kuin voimme sanoa mitään konkreettista tekoälyn ja Y-alustan suhteesta, on sovittava AI-palvelun määritelmästä.
- **Tekoälypalvelu** tässä puheenvuorossa tarkoittaa ensi sijassa loppukäyttäjälle (joka voi olla kansalainen, yritys tai viranomainen tai näiden yhdistelmä) kohdennettua sähköistä palvelua, jonka taustajärjestelmä soveltaa AI-tekniologiaa: datamassoja, suurta laskentakyvykkyyttä, oppivuutta ja sopeutumiskykyä.
- Tämän esityksen ajatuksia voidaan osin soveltaa myös ns. perinteisen ICT:n piirissä tuotettaviin palveluihin. Tarkka rajanveto ICT:n ja AI:n välillä on joka tapauksessa lähes mahdotonta, koska "se mikä oli eilen tekoälyä, onkin tänään vain digitalisaatiota".



Mistä puhumme, kun puhumme tekoälypalveluista? 2/2

- Toinen asia, josta on sovittava, liittyy siihen, millaisesta tekoälypalvelusta haluamme erityisesti puhua. Wikipedia liittää tekoälyn seuraaviin aiheisiin:
 - **Knowledge reasoning:** representing information about the world in a form that a computer system can utilize to solve complex tasks such as diagnosing a medical condition or having a dialog in a natural language.
 - **Planning:** realization of strategies or action sequences, typically for execution by intelligent agents, autonomous robots and unmanned vehicles
 - **Machine learning:** algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead
 - **Natural language processing:** interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.
 - **Computer vision:** deals with how computers can be made to gain high-level understanding from digital images or videos
 - **Robotics**
 - **Artificial general intelligence**
- Mitkä näistä osa-alueista ovat relevantteja Yhteentoimivuusalustan kannalta? Vasta kun osaamme vastata tähän, voimme alkaa varovasti hahmottaa, millaisia palveluja Y-alustan ja tekoälyn liitto voisi auttaa kehittämään.



Y-alusta on tekoälypalvelullekin vain pohja

- Yleisellä tasolla voi olettaa, että Y-alustaa voidaan käyttää tekoälytoteutusten **semanttisena pohjana**, jonka päälle voidaan ehkä mallintaa algoritmeja (joihin **ihmiset** määrittelevät parametrit, niiden arvot ja suhteet) joiden aktivoimiseksi tarvitaan riittävästi isoa dataa, joka on laatuvalidoitu niin eheyden, ajantasan kuin vinoumien suhteen
- AI ei käy kukkumaan eikä opi omia aikojaan, vaan ihmistyötä tarvitaan, alkuvaiheessa paljon.



Sanastot-työkalu

- Laadukkaan training datan yleisistä ominaisuuksista eheä data voi hyvin tarkoittaa myös **semanttisesti eheää dataa**: kun aineistossa käsitteiden ja muiden tietokomponenttien **merkitykset** on varmasti harmonisoitu, datan tietosisältö on ainakin siltä osin AI-kelpoista.
- Tulevien tietalueiden tuottamat määrittelevät, koneluettavat sanastot Y-alustalla antavat eväitä tuottaa sisällöllisesti yhtenäisiä usean toimijan palveluita, kun keskeiset käsiteavaruudet ovat yhteisiä. Sanastojen ansiosta AI:n tarvitsemaa luonnollisen kielen opetusdataa voi olla nykyistä selvästi helpompi muodostaa.
 - *Etiikkapointti*: pelkkä formaali ja semanttisen määrittelyn yhteentoimivuus ei estä bias-ongelmaa, jos aineisto jo lähtökohtaisesti heijastaa yhteiskunnan syviä vinoumia!
 - *Avoin kysymys*: jos Sanastot-aineistoja peilataan NLP:n avulla strukturoimattomiin tekstimassoihin, voivatko Sanastoissa olevat käsitteet saada koneellisesti lisää tietoa, esim. todennäköisiä synonyymeja? Jos, niiden käyttöarvo kasvaa niin järjestelmissä kuin ihmiskäytössä.



Tietomallit-työkalu

- Y-alustalla kukin tietomalli kertoo a) mitä yksittäinen elementti kuten luokka "huoneistolaji" ja sen attribuutti kuten "huoneistotunnus" sisältää, ja b) mistä luokista se koostuu ja miten luokat liittyvät mallissa toisiinsa → **yhdenmukaiset tietomallit mahdollistavat datan hyödyntämisen useasta eri lähteestä**
- Periaatteessa voitaisiin siis rakentaa **elämäntapahtumien tietomalleja** joissa on kuvattu määriteltyihin elämäntilanteisiin liittyvät toimijat, tarvittavat asiakirjat (esim. todistukset, luvat yms.) ja palvelut
 - AuroraAI:n suunnittelussa on jo tarkasteltu omaisen kuolema –elämäntapahtumaa ja käyty läpi, miten erilaiset palvelut voivat tehdä tarvittavat toimenpiteet kun puhuvat yhteistä kieltä
 - *Avoin kysymys*: pitäisikö kukin elämäntapahtumaan liittyvä **palvelu** kuvata Y-alustalle omana tietomallinaan, jolla on omat attribuuttinsa? Tulisiko elämäntapahtuman tietomallista silloin "tietomalleja sisältävä tietomalli?" Voisiko tekoäly tässä hyödyntää myös Palvelutietovarantoa, olettaen että sen sisältämä data on käyttökelpoista, riittävää ja asiallisesti ok?



Tekoälyn toiminta

- Tekoäly ei vielä pitkään aikaa tuota Y-tietomalleja. Määritellyt käsitteet, niistä koostuvat sanastot ja käsitteitä käyttävät tietomallit syntyvät edelleen ihmistyönä.
- Tekoälyn rooli Y-alustan aineistojen päällä operoivana toimijana olisi todennäköisesti välittäjä/yhdistäjä: tarjolla olevan tiedon analysointi ja tavoitteen kannalta relevanttien tietojen liittäminen toisiinsa tai kokoaminen muulla tavalla yhteen.
- Algoritmista ja palvelun tavoitteista riippuen AI voisi myös noutaa ja käyttää tilannekohtaisesti tarpeellisia tietoa muista julkisista aineistoista kuten esim. Finlexistä tai verkkosivustoilta, jälkimmäiset toki tiukalla laatu- ja luotettavuuskriteeristöllä tai jopa valmiiksi rajatun domain-valikoiman puitteissa (emme siis halua hallinnon AI:n noutavan aineistoa Vauva.fi:n keskusteluista, joista sen tosin voisi opettaa tekemään lauluja... jos siis joku ei niitä jo tekisi)
- **AI ei siis Y-alustan yhteydessä todennäköisesti loisi uutta aineistoa, mutta pystyisi aktivoimaan määriteltyjä tietomalleja palvelukäyttöön tunnistamalla, valikoimalla ja yhdistämällä tietoja, säilyttäen samalla merkityssisällöt**



Case Jonne

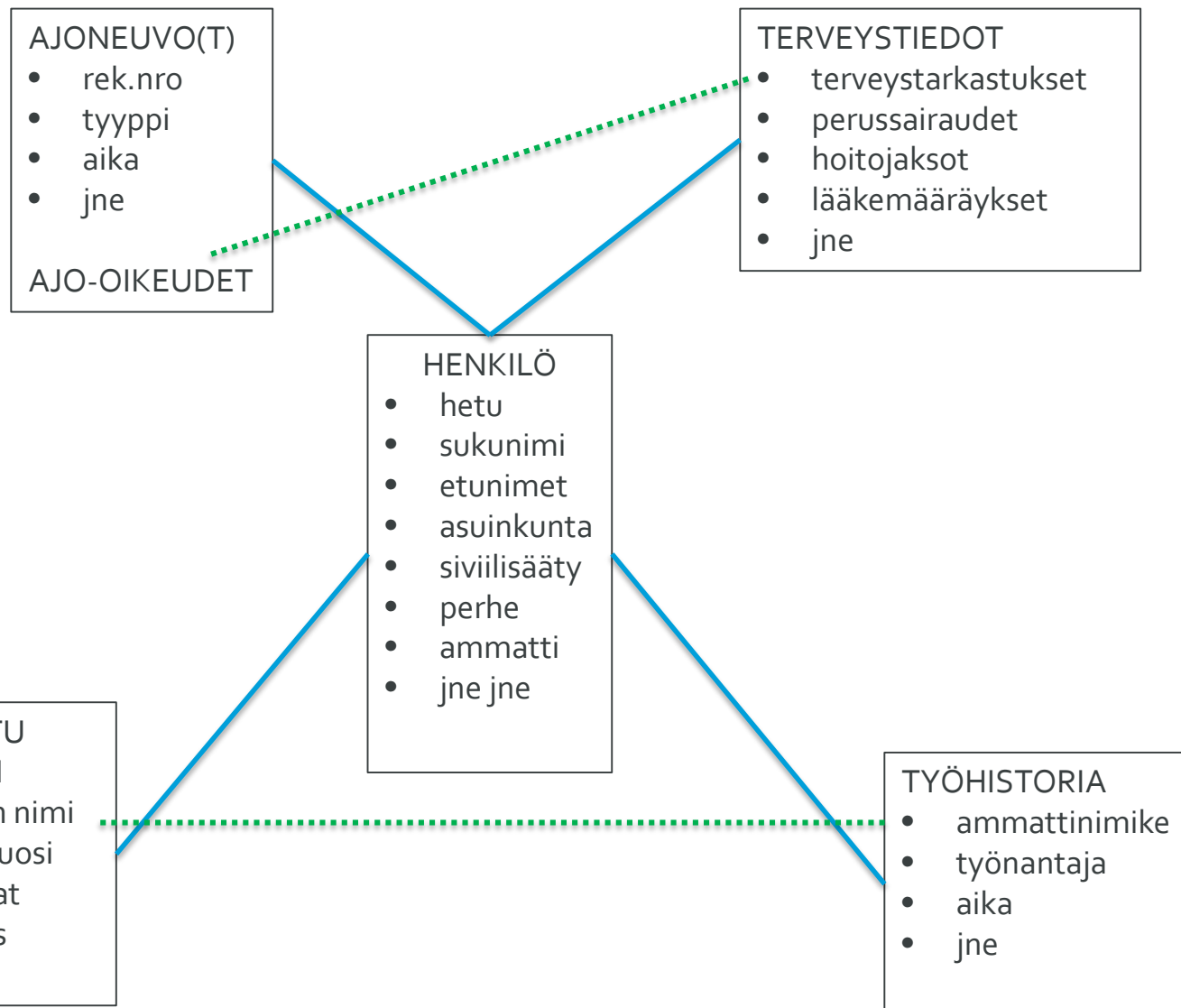
- Jonne on fiktiivinen 44-vuotias keskisuomalainen Jyväskylän seudulta. Jonnelta meni työt alta ja nyt on haku edessä.
- Jonne painelee alueensa TE-palvelujen toimipisteeseen ilmoittauduttuaan työttömäksi työnhakijaksi.
- Yhteentoimivuusalustalle on toteutettu työnhakijan tietomalli, jonka tarkoitus on koneellisesti koota TE-palveluille asiakkaasta riittävät tiedot.
- Tässä AI:n luomassa koosteessa yhdistetään useiden toimijoiden ja rekisterien tietoja asiakkaasta.
- Koosteen käyttäminen (ehkä jopa sen luonti) vaatii asianomaisen henkilön nimenomaisen suostumuksen.

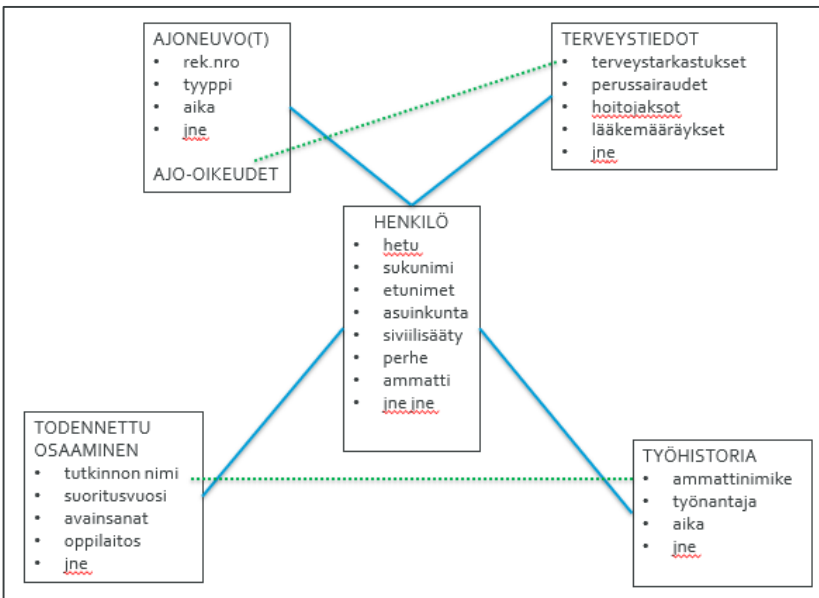


Työnhakijan tietomalli (seinästä revitty "kunhan tässä nyt jotain on" –esimerkki)

Jokaisen yksittäisen tiedon sisältö näissä kaikissa on määritelty Y-alustassa

Tiedot tulevat eri lähteistä, työhistoria myös sisäisesti eri puolilta (Tilastokeskus, PRH, Verö...)





JONNEN ALUEEN

- julkiset palvelut (PTV+webcrawler?)
- koulutuspalvelut (PTV+webcrawler?)
- aikuisopiskelijan tuet ja etuudet (Kela)
- avoimet työpaikat (Jonne-relevantit)
- työnantajat (Jonne-relevantit)
- julkinen liikenne (Jonnen koti – työ)

AI lähettää koosteen työvoimaneuvojalle mutta suljettuna!



TE-henkilö kertoo tapaamisen alussa että Jonnen asiaa voidaan tehokkaasti käsitellä häneen liittyvän tietopaketin avulla, mutta Jonnen pitää antaa sen avaamiseen lupa, koska kyse on hänen tiedoistaan.

Jonne päättää itse, antaako hän luvan tietojen käyttöön. Jos antaa, hänen täytyy vahvistaa suostumus vahvalla tunnistamisvälineellä.



Jotta totuus ei unohtuisi...

- Jonnea ei ole olemassa
- Vasta muutamat organisaatiot tässä vaiheessa osallistuvat Y-alustaan
- Eri viranomaisten tietojen kuvatuskaltaiselle yhdistämiselle on lainsäädännössä toistaiseksi esteitä; tiedoilla on rekisterikohtaiset käyttöperustevaatimukset → voisiko esim. työnhakuprosessin tehostaminen olla tulevaisuudessa käyttöperuste eräiden rekisterien tietojen käyttöön tässä kuvatuskaltaisella tavalla?
- Tässä fiktiivisessä Jonne-esimerkissä voi olla hyvinkin paljon asiavirheitä; se pyrkii vain **kuvaamaan mekanisme**, ei mitään todellista tietokokonaisuutta.



Kiitos!

<http://yhteentoimiva.suomi.fi>

yhteentoimivuus@vrk.fi

[Yhteentoimivuus.slack.com](https://yhteentoimivuus.slack.com)

marko.latvanen@vrk.fi

