

Heli Kautonen

## TIEDEKUSTANTAJAT OTTAVAT VAUHTIA TEKOÄLYYN

Mitä tiedejulkaisemisen käytäntöjä tekoäly haastaa ja mitä uusia mahdollisuuksia se avaa? Frankfurtin kirjamesset tarjoivat kirjastoammattilaiselle kiinnostavan tilaisuuden tutustua tekoälyvisioihin tiedekustantajien näkökulmasta.

37

Osallistuin Frankfurtin kirjamesseille lokakuun puolivälissä tarkoitukseni tutustua suomalaiseen kirjallisuusvientiin alan tärkeimmässä eurooppalaisessa tapahtumassa. Satuun eräänä messupäivänä kulkemaan tiedekustantajien hallissa juuri silloin, kun eräällä lavalta esiteltiin kustantajille tekoälysovelusten mahdollisuuksia ja haasteita. Kuuntelemaan jääminen kannatti, sillä nappasin kännykkääni monia uusia linkkivinkkejä ja kuulin kiinnostavia keskusteluja. Tässä artikkelissa raportoin kuulemaani ja näkemääni sekä taustoitan aihetta koskevilla lähteillä.

### Tiedekustantajat tekoälytoiveikkaita

**V**ielä kymmenisen vuotta sitten digitaalinen teknologia herätti kustannusalalla pelkoa siitä, että painettu kirja häviää markkinoilta ja liiketoimintamallit täytyy tyystin uudistaa. Tänä päivänä näytetään olevan vastaavan murroksen äärellä, kun tekoälyksi kutsutut teknologiset ratkaisut ovat ilmestymässä julkaisemisenkin toimialalle. Johtuneeko aiemmista digitalisaatiokokemuksista, että huolestumisen sijaan nämä uudet teknologiat herättävätkin kustantajissa

WHAT  
HAPPENS  
WHEN  
A  
MACHINE  
CURATES  
LITERATURE  
AND  
ART?

toiveikkuutta. Vasta julkaistun markkinakartoituksen mukaan monet kustantajat aikovat tulevan vuoden sisällä investoida tekoälysovelluksiin, elleivät ole sitä jo tehneet (AI in Academic Publishing Survey 2019).

Uudet teknologiat voivat näyttäytyä houkuttelevina myös siksi, että kustannusalan ympärillä on muuta kuohuntaa. Kaunokirjallisuuden kustantajat sopeutuvat viihdekulututtamisen uusiin muotoihin kasvattamalla tilausperustaisia e-kirja- ja äänikirjapalveluitaan (Albanese 2019). Tällaiset viihdekanavapalvelut hyödyntävät suosittelualgoritmeja, jotka tarjoavat kullekin käyttäjälle juuri häntä houkuttelevaa sisältöä.

Tiedekustantamisen puolella liiketoiminnan perustuksia on järjestyttänyt tiedeyhteisöissä voimistunut avoimen julkaisemisen liike, jonka tämän hetken vahvin vaikuttaja on eurooppalainen Plan S -aloite. Kustantajien täytyy nyt sopeutua nopeasti tieteen rahoittajilta tullessiin vaatimuksiin ja samalla etsiä keinoja kilpailukykyä säilyttämiseen. Kaikkein, mikä mahdollistaisi tiedekustantamon olemassaolon Plan S:n voimaantumisen jälkeen näytetään suhtautuvan nopeasti (vrt. mm. Albanese 2019, Sayer 2019).

Kustantajille olennaista uusien teknologisten innovaatioiden käyttöönotossa on oletettu hyöty. Yrityksellä täytyy olla kirkas kuva siitä, mitä investoinnilla saavutetaan, miten paljon se tuo säästöä ja miten uudet teknologiat istuvat perinteisiin toimituksen ja

tuotannon prosesseihin (Paneelikeskustelu 18.10.2019, AI in Academic Publishing Survey 2019).

## Tehokkuutta työprosesseihin

**K**ustannusalan kilpailu on kovaa ja siksi kilpailuetua etsitään kaikista mahdollisista tuotannon vaiheista. Työvoimavaltaisia prosesseja mallinnetaan ja opetetaan tietokoneille, toiveena poistaa toimitustyöstä turhat pullonkaulat ja inhimilliset virheet (Groth 2019). Kustantajat pitävät työajan säästöä tällä hetkellä tärkeimpänä tekoälysovellusten tarjoamana hyötynä (AI in Academic Publishing Survey 2019).

Lähetettyjen käsikirjoitusten hitaaksi koettuun käsittelyyn on lähdetty kehittämään työkaluja, joiden avulla julkaisija voi plagiaattitarkistuksen ohella tarkistaa käsikirjoituksen teknisen virheetömyyden, kielen ja tärkeimmät väittämät tai arvioida tutkimuksen uutuusarvoa. Useammatkin toimittajat tarjoavat tällaisia palveluja, joka hyödyntävät laajoja julkaistujen artikkeleiden tietokantoja, koneoppimista ja kieliteknologioita. Frankfurtissa esillä olivat ainakin Cenevo (<https://www.ceneopublisherservices.com/editorial-journals>), Impelsys (<https://www.impelsys.com/publishing-solutions>) ja Unsilo (<https://unsilo.ai/unsilo-evaluate>). Unsilon palvelussa on lisäksi käsikirjoituksen aihepiiriin sopivien vertaisarvioijien etsintä.

Frankfurtissa näin painettuna versiona Springerin tänä keväänä julkai-

seman maailman ensimmäisen koneen koostaman tieteellisen julkaisun (lähdeluettelossa: Writer 2019). Tällä kokeilullaan Springer on kurottanut askeleen pidemmälle ja kokeillut, miten tämänhetkiseltyä tekoälyltä sujuu tieteellisen tekstin tuottaminen.

Kustantajien keskusteluissa nousi esille toiveita myös tutkimustrendien automaattiseen tunnistamiseen, sillä trendikkäimmät aiheet lisäävät latauksia ja nostavat julkaisun vaikutavuuskerrointa. Muiden muassa kirjastoillekin tuotteitaan markkinoivan Yewnon valikoimassa on palvelu, jolla voi tarkkailla, millaiset tutkimuksen teemat ovat nousussa tai laskussa (ks. <https://www.yewno.com/unearth>).

### Parempia palveluja ja käyttöliittymiä

**T**utkimustiedon parempi saatuus on keskeinen tavoite niin kirjastoille kuin kustantajillekin. Kun tieteellisten julkaisujen määrä kasvaa vuosittain muutamalla miljoonalla (Johnson ym. 2018), ajankohtaisen tai vanhemmankin tutkimuksen löytäminen muuttuu perinteisin tiedonhakukeinoin sängen työlääksi. Uudenlaisia apuvälineitä siis tarvitaan.

Kuten kirjastot niin myös kustantajat pyrkivät edistämään aineistojen löydettävyyttä metadatan avulla, kontrolloiduilla ja linkitetyillä sanastoilla. Markkinaselvitykseen vastanneista kustantajista lähes puolet hyödynsivät tekoälyä automaattiseen tekstien analysointiin ja metadatan tuottamiseen (AI

in Academic Publishing Survey 2019). Vaikka metadatuotanto katsotaan edelleen kustannustalon sisäiseksi toiminnoksi, myös luotettavien toimittajien palveluita arvostetaan. Esimerkkinä Frankfurtissa mainittiin 1970-luvulta saakka toiminut yhdysvaltalainen Access Innovations, joka lupaa mm. tuottaa joka organisaation tarpeisiin sopivan sanaston tai ontologian (ks. <https://www.accessinn.com/services>).

Algoritmiiketoiminta näyttää kasvattavan osuuttaan myös tiedejulkaisemisen ympärillä. Markkinamenestystä haetaan panostamalla parempiin käyttöliittymiin ja hakutulosten visualisointeihin, kuten ovat tehneet Frankfurtin lavalla esimerkeiksi nostetut Yewno (<https://www.yewno.com/discover>) ja Semantic Scholar (<https://www.semanticscholar.org>) (Groth ym. 2019).

Lisääntyvästä ohjelmisto- ja palvelutarjonnasta huolimatta valtaosa kustantajista ymmärtää, että tekoälyn hyödyntäminen edellyttää myös oman organisaation osaamisen kehittämistä (AI in Academic Publishing Survey 2019). Kustantajat kannustavat toisiaan investoimaan dataan ja sen älykkäiseen käsittelyyn (Tan 2019). Kaiken tämän kuulemani ja lukemani perusteella meneillään näyttää olevan tiedekustantajien vauhdinotto tekoälyyn.

### Entä vaikutukset?

Vierailu Frankfurtin kirjamessuille osoitti minulle sen, että kirjastojen on syytä seurata tiedekustantajien tekoälytoimia. Kuulemissani puheen-



vuoroissa korostettiin, että kustantajat pitävät laatu- ja näkökulmia erittäin tärkeinä (Groth ym. 2019, Prabhu 2019, vrt. Shoenenberger 2019). Silti markkinaselvitys viittaa siihen, että harva kustantaja panostaa tällä hetkellä uusien sovellustensa mahdollisten vinoumien tarkastamiseen, saatikka ehkäisyyn (AI in Academic Publishing Survey 2019).

Matemaattisten mallien tuottama tieto on lähtökohtaisesti erilaista kuin ihmisen tuottama. Algoritmien tai datan aiheuttamat vinoumat voivat johtaa dramaattisesti virheellisiin päätelmiin tai vain huonoon lopputulokseen (ks. esim. Lee ym. 2019). Millaiseen suuntaan tiede mahtaa kääntyä, jos tai kun algoritmien annetaan valita tieteellisten artikkeleiden vertaisarvioijat tai arvioida käsikirjoitusten tutkimuksellinen uutuusarvo? Pitäisikö tiedeyhteisön jollain tavoin säädellä, miten paljon koneiden sallitaan avustaa tieteellisen tekstin tuottamista tai parantamista?

Tiedekustantajien siirtymistä käyttämään tekoälysovelluksia voidaan tarkastella myös tieteellisten kirjastojen edun näkökulmasta. Miten mahtavat tehokkaammat tuotantoprosessit näkyä lissensiosopimusten hintojen tai avoimen julkaisemisen maksujen perusteissa?

### **Pysykäämme kirjastoissa kuulolla ja osallistukaamme keskusteluun**

Vastustanko sitä, että kustantajat ottavat tekoälyteknologiaa käyttöönsä? Päinvastoin. Toivon, että uudet väli-

neet ylläpitävät julkaisualan monimuotoisuutta tarjoamalla pienemmille ja voittoa tuottamattomillekin kustantajillekin keinoja ylläpitää toimintaansa. Koska tiedekirjastot ja tiedekustantajat palvelevat samaa asiakaskuntaa, meidän tulisi käydä avointa keskustelua tekoälyteknologiaa hyödyntävien järjestelmien positiivisista ja negatiivisista vaikutuksista tieteen tekemisen käytäntöihin ja rakenteisiin.

Tieteellisen tiedon laadun takaaaminen edellyttää jatkossa lisää läpinäkyvyyttä siihen, missä julkaisemisen prosesseissa algoritmeja hyödynnetään ja miten. Väitän, että tieteellisten kirjastojen tulisi olla mukana arvioimassa julkaisualalla käytettävien tietojärjestelmien vääristymävaikutuksia (*bias impact statements*, vrt. Lee ym. 2019). Emme myöskään saisi unohtaa omien, algoritmeja hyödyntävien järjestelmien kriittistä arviointia ja läpinäkyvyyttä. (Toisin sanoen myös mainiossa Annif.org:ssa tulisi olla kuvaus mahdollisista vääristymävaikutuksista, eikä vain!)

Tieteen tekemisen ja tiedeyhteisöjen edun mukaista olisi, että tiedekustantajat, ohjelmistotoimittajat ja tieteelliset kirjastot löytäisivät yhteisiä foorumeita, missä puntaroida uusien teknologioiden tarjoamia uhkia ja mahdollisuuksia. Mitä jos meitä kirjastolaisia olisi ensi vuonna Frankfurtissa muutama muukin kuuntelemassa – ja osallistumassa keskusteluun tiedekustantajien kanssa. ♦

## Koetut lähteet:

GROTH, MICHAEL; LONGO, JIM; SCHOENENBERGER, HENNING & THOMAS, NIELS PETER. AI 2.0 Machine-Generated Content, Intelligent Automation, and the Future of Academic Publishing. Paneelikeskustelu 18.10.2019. Frankfurtin kirjamesut, Saksa.

PRABHU, JOHN. Leveraging Technology to Create Intelligent Content. (Accessibility, Metadata, Indexing & Abstracting). Esitys 18.10.2019. Frankfurtin kirjamesut, Saksa.

## Julkaistut lähteet:

AI in Academic Publishing Survey. 2019. Unisilo.ai. <https://unisilo.ai/wp-content/uploads/2019/11/unisilo-Survey-on-ai-in-Academic-Publishing-2019.pdf> (Luettu 23.11.2019)

ALBANESE, ANDREW RICHARD 2019: What a Difference a Decade Makes. Frankfurt Book Fair - The Digital Spotlight, 3–4. Publishers Weekly.

GROTH, MICHAEL 2019: ai Goes from Disruptive to Imperative. Frankfurt Book Fair - The Digital Spotlight, 10–12. Publishers Weekly.

JOHNSON, ROB; WATKINSON, ANTHONY & MABE, MICHAEL 2018. The STM Report, Fifth edition. An overview of scientific and scholarly publishing. stm: International Association of Scientific, Technical and Medical Publishers. [https://www.stm-assoc.org/2018\\_10\\_04\\_STM\\_Report\\_2018.pdf](https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf) (Luettu 20.11.2019)

LEE, NICOL TURNER; RESNICK, PAUL & BARTON, GENIE 2019. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Report. Brookings. <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/> (Luettu 23.11.2019)

SAYER, LIZZIE 2019: Scholarly publishers also need a more consistent approach. Interview with Steven Inchcoombe of Springer Nature on Plan S and Open Access. Blog post 20.3.2019. International Science Council. <https://council.science/current/blog/scholarly-publishers-also-need-a-more-consistent-approach-interview-with-steven-inchcoombe-of-springer-nature-on-plan-s-and-open-access> (Luettu 23.11.2019)

SCHOENENBERGER, HENNING (2019). Preface. Teoksessa Beta Writer: Lithium-Ion Batteries. A Machine-Generated Summary of Current Research, ss. v–xxiii. Springer Nature. <https://doi.org/10.1007/978-3-030-16800-1>

TAN, TERI (2019). Fast or Slow, Digitization Does Go on. Frankfurt Book Fair - The Digital Spotlight, 16–22. Publishers Weekly.

WRITER, BETA toim. (2019). Lithium-Ion Batteries. A Machine-Generated Summary of Current Research. Springer Nature. <https://doi.org/10.1007/978-3-030-16800-1>

## Tietoa kirjoittajasta

HELI KAUTONEN

Suomalaisen Kirjallisuuden Seura

[heli.kautonen@stks.fi](mailto:heli.kautonen@stks.fi)

<https://orcid.org/0000-0001-6652-1165>