

Tekoälyohjelma Iris.AI:n alkutaival, käyttäjäkokemukset ja uudet kilpailijat

Tekoälyohjelma Iris.ai otettiin Helsingin yliopiston kirjastossa käyttöön syksyllä 2017 ja Itä-Suomen yliopiston kirjastossa huhtikuussa 2018. Palvelua markkinoitiin tutkijan assistenttina, jolla tieteellisen tiedon määrä saadaan nopeammin haltuun. Millaisia olivat käyttäjien kokemukset Iris.aista?

”Iris AI on keinoälypohjainen työkalu open access -tutkimuksen kartoittamiseen. Tee haku millä tahansa tutkimuspaperin verkko-osoitteella, Iris lukee sen abstraktin ja rakentaa lukemansa pohjalta tutkimuskartan etsimällä siihen keskeisimmät sisällöt 30 miljoonan avoimen tutkimuspaperin joukosta.”

Näin määritteli nuori startup-yritys Iris.ai palvelutuotteensa syksyllä 2016, kun sitä tarjottiin ensimmäisen kerran Helsingin yliopiston kirjastolle. Yritys oli perustettu edellisenä vuonna 2015. Ensimmäinen yhteistyön vuosi yrityksen ja kirjaston kesken käytettiin palvelun kehittämiseen ja erityisesti keskityttiin siihen, millaista tiedonhakupalvelua tieteelliset kirjastot ovat tottuneet vaatimaan.

Edelläkävijäasiakkaana toimiminen on tuonut Helsingin yliopiston kirjastolle myös huomattavaa näkyvyyttä tiedonhaun kentällä toimivien organisaatioiden parissa. Paitsi suomalaisista organisaatioista niin myös useista eurooppalaisista kirjastoista on otettu yhteyttä ja toivottu kertomaan palvelun käytöstä ja sen vaikutuksista tiedonhakuun. Näin ollen kirjaston on ollut mahdollista profiloitua uusiin

palveluihin positiivisesti ja rohkeasti suhtautuvana proaktiivisena toimijana, aivan kuten suuren tieteellisen kirjaston tuleekin toimia.

Premium-versio ja sen Focus-työkalu valmistuivat syksyllä 2017 ja tässä vaiheessa Iris.ai (the.iris.ai/) tuli laajemminkin tieteellisten kirjastojen tietoisuuteen. Helsingin yliopiston kirjasto päätyi hankkimaan käyttölisenssin huhtikuussa 2018, Itä-Suomen yliopiston (UEF) kirjasto lokakuussa 2018. Lisäksi Iris.ai on ollut pilottikäytössä useissa yliopisto- ja ammattikorkeakoulukirjastoissa, ja palvelua on kehitetty paljon kirjastolaisten ja kirjaston asiakkaiden kommenttien pohjalta.

Miksi tekoäly tuli tieteelliseen tiedonhakuun?

On arvioitu, että joka päivä julkaistaan noin 4 000 uutta tieteellistä artikkelia tai muuta julkaisua. Vuositasolla puhutaan siis puolestatoista miljoonasta tuotoksesta. Suuret monialaiset tietoaaineistot ja niiden luonnollisen kielen parempi ymmärtäminen haastavat perinteiset hakusanoilla ja tietokantakohtaisilla asiasanastoilla toimivat tietokannat. Iris.ai:n taustalla

UPDATE

true

MORE

false

VALUE=1
+trackNo



Kuva 1. Iris.ai käsittekartta aiheeseen E-leadership in the IT industry and in cross-cultural global setting

olevassa CORE-tietokannassa on tällä hetkellä yli 200 miljoona dokumenttia, kun Iris.ai aloittaessa niitä oli noin 30 miljoonaa. PubMed-tietokantaan lisättiin 192 057 uutta tietuetta vuonna 1967, vuonna 2020 määrä oli jo yli 1,4 miljoonaa.

Ovatko suuret julkaisumäärät tutkijan hallittavissa? Tekoälyn on tarkoitus säästää tutkijan aikaa.

Julkaisumäärän lisäksi kasvavat vaatimukset tieteidenvälisestä tutkimuksesta ja sen avulla saatavista innovaatioista ovat lisänneet paineita etsiä uusia ideoita tieteellisen tiedon kokonaisuudesta, eikä ainoastaan yhden tieteenalan tietokannoista.

Samaan aikaan keskusteluun on noussut myös avoin tiede ja tutkimusjulkaisujen entistä avoimempi

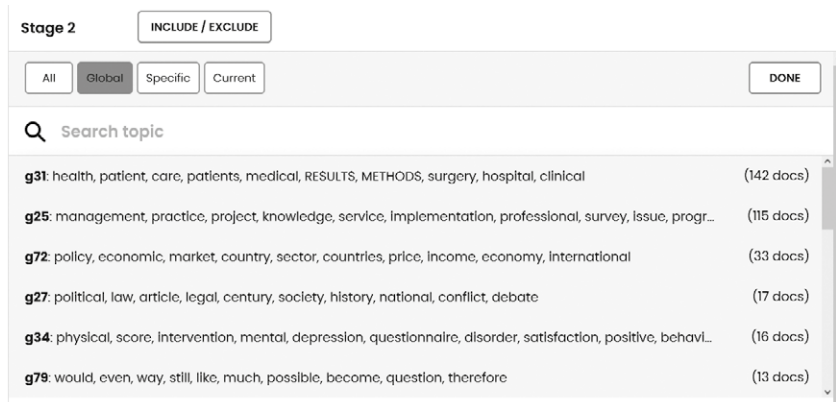
saatavuus ja tiedekustantamisen ja -julkaisemisen murros. UEF-kirjastossa pohdittiin myös, olisiko Iris.ai mahdollisesti yksi toimiva väline avointen julkaisujen hakuun.

Iris.ai hakuvälineenä

Iris.ai ei markkinoinut itseään tietokantana kirjastoille, vaan tutkijan assistenttina, jolla tieteellisen tiedon määrä saadaan nopeammin haltuun. Se ikään kuin ”lukee” dokumentit ja seuroi käyttäjälleen niistä relevanteimmat.

Iris.ai mahdollistaa tiedonhaun kiinnostavan tutkimusartikkelin URL-tai DOI-linkkiin pohjautuen ja näyttää hakutulokset visuaalisena käsittekarttana (Kuva 1). Maksullinen organisaatiolisenssi mahdollistaa karttojen muokkaamiseen ja haun tarkentamisen Exploring- ja Focusing-työkaluilla. Innovatiivista oli yksittäisten hakusanojen ja Boolean logiikan sijaan hakumahdollisuus omin sanoin 300–500 sanan mittaisella tiivistelmällä ja tulosten visuaalinen esittäminen.

Moni jättää haun käsittekarttavaiheeseen, vaikka Focusing-työkalulla on tarkoitus tarkentaa tulosta, ottaa mukaan ja sulkea pois käsitteitä ja käsiteryypäitä.



Kuva 2. Iris.ai:n focusing-työkalun käsityrväs.

Käyttäjien kokemuksia – ruusuja

Itä-Suomen yliopiston kirjastossa on kerätty palautetta Iris.ain käytöstä opiskelijoilta. Seuraavassa esitetyt näkemykset ja sitaatit ovat Maarit Putouksen vuosina 2020–2021 vetämiltä tiedonhaun Moodle-kursseilta, joissa opiskelijoilla on ollut ylimääräisenä ja vapaaehtoisena tehtävänä kokeilla Iris. ai:ta. Kokeilijoita on ollut parikymmentä. Lisäksi palautetta on kerätty muutamilta tutkijoilta, joiden kanssa Iris.aita on käytetty heidän tutkimusongelmiinsa liittyvissä tiedonhauissa.

Näin positiivisesti Iris.aita kuvailee eräs käyttäjä:

”Auttaa löytämään sellaisia uusia yhtäläisyyksiä mitä ei välttämättä itse tajuaisi etsiä. Toki samaa arvoa saa Googlen algoritmin avulla mutta olen kokenut molempien ikään kuin täydentävän toisiaan ja molempia käyttämällä löytää mainiosti uutta relevanttia luettavaa. On se helppoa näin nykypäivänä!”

Muita Iris.ain myönteisiksi arvioituja ominaisuuksia ovat:

- nopeus
- luonnollisella kielellä haku
- sanojen yhteyksiä ja kategorisointeja/ryhmittelyjä
- vastaan voi tulla sellaisiakin käsitteitä ja yhteyksiä, joita ei ole edes osannut ajatella
- auttaa tutkimuksen alkuvaiheessa hahmottamaan aiheen taustaa

- hieman laajemman perspektiivin/ kontekstin saaminen
- esim. opiskelijalle, joka ei tunne vielä käsitteitä eikä osaa hakusanoilla hakuajan rajat

Visuaalisuudesta – käsitekartasta – tykkäävät kaikki!

”Käyttöliittymä on käytettävyydeltään virkistävän moderni, ja visuaalisesti pelkistetty ja kaunis verrattuna moneen muuhun käyttämäni hakutietokantaan. Hakutulokset oli melko pirstaleinen ja artikkelien tuoreus vaihteli melkoisesti. Vanhin oli vuodelta 1994 ja uusin vuodelta 2016. Oli mielenkiintoista kokeilla tämän tyyppistä käyttöliittymää.”

Käyttäjien kokemuksia – risuja

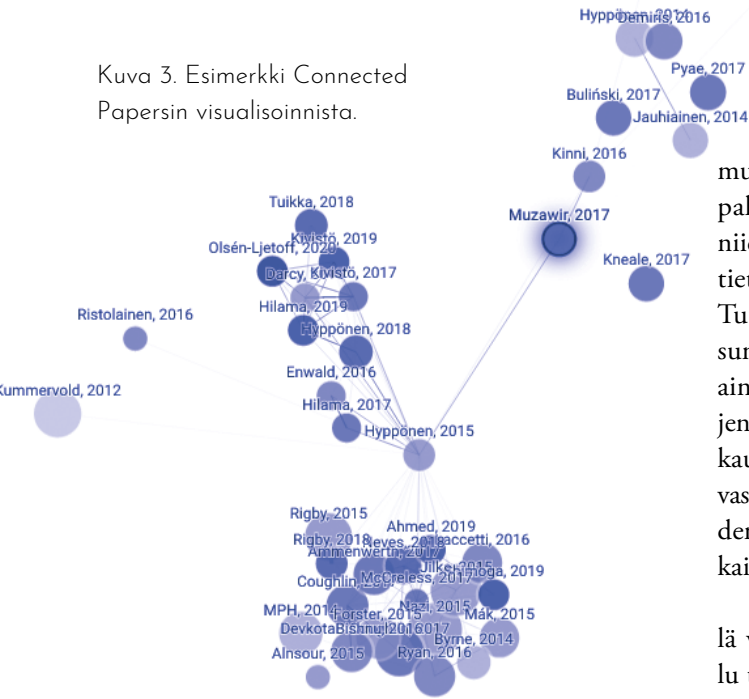
Iris.ain käyttöön liittyy myös kielteisiä kokemuksia.

”Eiväthän nämä ole oikeaa asiaa! Miten tähän voi luottaa? Tähän ei kannata tuhata aikaa! Google Scholar pystyy paljon parempaan.”

Ensimmäinen ongelma on, että käsitekartan sisältö on ”mitä sattuu”. Tulokset eivät näytä relevanteilta, tärkeimmät käsitteet puuttuvat kartalta eivätkä asiakokonaisuudet yhdisty keskenään. Toisaalta tutkimusongelmissa, joissa on tarkat käsitteet, myös tekoäly toimii paremmin.



Kuva 3. Esimerkki Connected Papersin visualisoinnista.



Tulosmäärät nähtiin joskus yllättävän pieninä verrattuna perinteisiin hakumenetelmiin. Mitä, miksi ja miten jättää tuloksista pois?

”Minulle jäi vielä epäselväksi, että miten saan tämän ohjelman avulla varmistettua, että tekstit ovat tieteellisiä? Onko sellaista rajausmahdollisuutta edes olemassa tässä ohjelmassa?”

Pohjadataan merkitys tieteellisessä tiedonhauissa on olennainen. CORE-tietokantaan haravoidaan eri organisaatioiden julkaisuarkistoja, ja ne pitävät sisällään myös gradu-, kandidaatintyö- ja amk-opinnäytetyötasoisia julkaisuja. Vertaisarviointirajausta tai opinnäytetöiden pois sulkua ei ole.

PubMed/Medline on integroituna Iriksen, mutta tämä ei riitä. Monet

muut open access -aineistoa hakevat palvelut ovat maksutta käytössä ja niiden vaikuttimena näyttäisi olevan tieteen avoimuus ja saavutettavuus. Tutkimustiedosta $\frac{2}{3}$ on vielä maksumuurien takana ja suuri osa tästä aineistosta on käytettävissä kirjastojen perinteisten väylien/tietokantojen kautta. Aineiston integrointi maksua vastaan Iris.ai:n haettavaksi ei ole muiden kuin Iris.ain näkökulmasta ratkaisu.

Kielen ymmärtämisessä on Iriksellä vielä oppimista. Fokusointityökalu tekee esim. tällaisia käsiteryypäitä: would, even, way, still, like, much, possible, become, question, therefore.

Harva palautetta kirjoittanut käyttäjä on edennyt Focusing-työkalulla. Karttavaiheessa käyttäjät innostuvat – tai turhautuvat. Focusing-työkalu on vaikeasti hahmotettava ja työläs. Jos viitetulos on pieni, hakutuloksen/viitelistan silmäilee nopeasti.

Uusia yrittäjiä tekoälypohjaisen tiedonhakuun

Tieteellisen tiedon valtavan kasvun myötä tiedonhaun markkinoille tulee pioneiryritysten jälkeen jatkuvasti toimijoita. Toisena syynä on tietenkin myös tekoälyn voimakas kehittyminen ja siihen kohdistuvat panostukset. Avoin julkaiseminen lisää massaa uusien tulokkaiden käyttämiin tietokantoihin (esim. Semantic Scholar), jolloin tekoälystä saadaan mahdollisimman paljon irti.

Tässä kolme esimerkkiä tekoälyä hyödyntävistä palveluista, joissa kaikissa tulokset esitetään visuaalisena karttana ja usein haun lähtökohtana on kiinnostava artikkeli (vrt. helmenkasvatus- tai lumipallostrategia):

- **Connected Papers:** Visuaalinen työkalu, jonka algoritmi pohjaa artikkelien samankaltaisuuteen, päällekkäisiin viittauksiin ja bibliografiseen kytkentään (Kuva 3.). Palvelu on maksuttomasti käytettävissä ja sitä hyödynnetään mm. Helsingin yliopiston Meilahden lääketieteellisen kampuksen uusien artikkelien twiittaamisessa. (www.connectedpapers.com/)

- **Open Knowledge Maps:** Algoritmi etsii tekstin samankaltaisuuksia ja muodostaa niistä ryhmiä, joissa julkaisut muodostavat yhtenäisen kokonaisuuden. (openknowledgemaps.org/index)

- **ResearchRabbit:** Tuore tulokas tekoälypohjaiseen tiedonhaussa. Tässäkin pohjalla ovat käyttäjälle tärkeät artikkelit, joiden perusteella palvelu suosittelee muita julkaisuja. Se nimittääkin itseään "Spotify for papers" -iskulauseella. Palvelu eroaa tietokantojen "Find related papers" -optiosta. Niissä on useimmiten käytetty yksinkertaisesti tietokannan metadataa ja indeksoituja termejä. ResearchRabbit hyödyntää tekoälyä ja algoritmi pohjautuu suosittelujärjestelmään. (www.researchrabbit.ai/)

B==C

true

ONLINE

Tietoympäristö muuttuu

Kirjastojen tietoasiantuntijoiden näkökulmasta tekoäly aiheuttaa myös ongelmia. Kuten tunnettua, kaikissa tekoälyä käyttävissä palveluissa algoritmien toiminta ei ole läpinäkyvää, johtuen erityisesti siitä, että ne ovat palvelujen ydintä ja liiketoiminnan perusta. Siten ei voida tietää miten hakutulos muodostuu ja esimerkiksi systemaattisissa kirjallisuuskatsauksissa niiden toistaminen on ongelmallista, sillä tekoäly ei kerro hakustrategiaansa.

Tieteellisissä kirjastoissa tarvitaan tietoisuutta ja ymmärrystä algoritmien toiminnasta tiedonhaussa. Tietoasiantuntijan on osattava valita ja ohjata tarkoituksenmukaisimpiin hakuvälineisiin. Osa asiantuntijuutta on verrata ja ottaa kantaa myös tekoälypohjaisiin järjestelmiin. ❖

KIRSI SALMI

Itä-Suomen yliopiston kirjasto
kirsi.salmi@uef.fi

JUKKA ENGLUND

Helsingin yliopiston kirjasto
jukka.englund@helsinki.fi

MAARIT PUTOUS

Itä-Suomen yliopiston kirjasto
maarit.putous@uef.fi

false