

Usein kysytyt kysymykset bibliometriikasta

www.aka.fi/tieteentila

Sisällys

Aineisto	3
Miten hyvin Web of Science -pohjainen aineisto kattaa eri alojen julkaisuja?	3
Miksi julkaisujen osoitteet pitää tarkistaa?	3
Miten aineistossa käsitellään yliopistojen yhteisten tutkimuskeskusten julkaisut?	3
Miten bibliometrisissä analyyseissä käytetty tieteenalaluokitus on tehty?	3
Viittausindikaattorit	4
Mikä on viittausindikaattori?	4
Miksi viittausmäärä normalisoidaan?	4
Miksi ja miten julkaisut ositetaan?	5
Mitä normalisoitu viittausindeksi mittaa ja miten se lasketaan?	5
Mitä top 10 -indeksi mittaa ja miten se lasketaan?	6
Miten pieni julkaisumäärä vaikuttaa top 10 -indeksiin?	6

Kysymys- ja vastauslistaa päivitetään tarpeen mukaan.

Lisätietoja:

Anu Nuutinen

Suunnittelu ja johdon tuki -yksikkö

Suomen Akatemia

sähköposti: [tieteentila\(at\)aka.fi](mailto:tieteentila(at)aka.fi)

www.aka.fi/tieteentila

Aineisto

Miten hyvin Web of Science -pohjainen aineisto kattaa eri alojen julkaisuja?

Julkaisukäytännöt ovat osin erilaisia eri tieteenaloilla. Tieteenalojen julkaisumäärät eivät ole kaikilta osin vertailukelpoisia keskenään, koska kansainväliset viitetietokannat eivät kata kaikkien tieteenalojen julkaisuja samalla tavalla. Web of Science -pohjainen aineisto (WoS) ei sisällä esimerkiksi tieteellisiä artikkeleita kokoomateoksissa tai kustannettuja tieteellisiä erillisteoksia. Kansainvälisten viitetietokantojen aineisto ei sovellu julkaisutoiminnan yksityiskohtaisempaan tarkasteluun yhteiskuntatieteellisillä ja humanistisilla aloilla yhtä hyvin kuin useilla muilla tieteenaloilla. Myös vertaisarvioitujen konferenssijulkaisujen osalta aineisto on puutteellinen, mikä vaikuttaa selkeästi esimerkiksi tietojenkäsittelytieteen julkaisumääriin.

WoS-aineiston kattavuutta voidaan arvioida laskemalla ns. tieteenalan sisäinen kattavuus eli se, kuinka suuri osa tieteenalan viittauksista osoittaa tietokantaan, jota analyysiin käytetään. Kliinisissä lääketieteissä sisäinen kattavuus on nykyisin 95 prosenttia, humanistisilla aloilla 51 prosenttia. Mitä suurempi on alan sisäinen kattavuus, sitä paremmin WoS-aineisto soveltuu kyseisen tieteenalan tieteellisen vaikuttavuuden analysoimiseen.

Miksi julkaisujen osoitteet pitää tarkistaa?

Organisaationimissä esiintyy useita erilaisia kirjoitusasuja. Julkaisujen osoitetiedoissa esiintyy myös suoranaisia virheitä. Jotta kaikki tietyn organisaation julkaisut tulevat lasketuksi tämän organisaation tilastoihin, Suomen tutkimusorganisaatioiden osoitetiedot on tarkastettu CSC:ssä.

Miten aineistossa käsitellään yliopistojen yhteisten tutkimuskeskusten julkaisut?

Yliopistojen yhteisiä tutkimuskeskuksia ovat esimerkiksi Helsinki Institute for Information Technology (HIIT), Helsinki Institute of Physics (HIP), FIMM Institute for Molecular Medicine Finland ja Turun PET-keskus. Näiden ja vastaavien tutkimuskeskusten julkaisuja on käsitelty seuraavasti (esimerkkinä HIIT):

- jos osoitekentässä on selkeästi ollut HIIT, Aalto University, julkaisupiste lasketaan vain Aalto-yliopistolle
- jos osoitekentässä on HIIT, Aalto University, University of Helsinki, molemmille 0,5 julkaisupistettä
- jos osoitekentässä on pelkästään HIIT eli ei voida tietää, kumman organisaation tutkijat ovat julkaisun tekijöitä, julkaisu sijoitetaan luokkaan muut eli se ei tuota pisteitä yliopistojen tilastoihin. Tällaisten julkaisujen, joita ei voida osoittaa yliopistoille, osuus on pääsääntöisesti pieni.

Miten bibliometrisissä analyyseissä käytetty tieteenalaluokitus on tehty?

Thomson Reuters luokittelee kaikki lehdet yhdelle tai useammalle alalle (subject category), joita on noin 250. Lehden numero voidaan luokitella 1–6 alalle. WoS-pohjaisessa aineistossa julkaisun on katsottu edustavan niitä tieteenaloja, joille Thomson Reuters on luokitellut lehden, jossa julkaisu on ilmestynyt.

Tieteen tila -hankkeen analyysien varten Thomson Reutersin käyttämät alat (subject category) on luokiteltu laajemmiksi tieteenaloiksi kansallisen tieteenalaluokituksen (2010) mukaan (ks. tarkemmin www.aka.fi/tieteentila > Menetelmät ja luokitukset > Bibliometristen analyysien tieteenalaluokitus). Kansallisessa luokituksessa on muutama tieteenala, joita vastaavia aloja ei sisälly Thomson Reutersin luokitukseen. Lisäksi joitakin kansallisen tieteenalaluokituksen aloja on yhdistetty tarkastelun kannalta suuremmiksi kokonaisuuksiksi. WoS-aineiston puutteellisen kattavuuden takia yhteiskuntatieteistä voidaan tarkastella psykologiaa ja taloustieteitä tieteenaloittain. Humanistiset tieteet eivät sisälly tieteenalakohtaiseen bibliometriseen analyysiin.

Viittausindikaattorit

Mikä on viittausindikaattori?

Viittausindikaattori on yleisnimitys viittausanalyysissä käytetyille bibliometrisille mittareille. Tieteen tila -hankkeen viittausanalyysit perustuvat Thomson Reutersin Web of Science -tietokantaan indeksoitujen tieteellisten lehtien julkaisuihin ja niiden saamiin viittauksiin. Käytetty aineisto kattaa vuosien 2000–2012 julkaisut julkaisutyypeissä article, letter ja review. Viittaukset on laskettu avoimen viittausikkunan perusteella. Itseviittaukset (tutkija viittaa aiemmin julkaisemaansa omaan julkaisuun) on poistettu, koska tavoitteena on analysoida ensisijaisesti julkaisujen tieteellistä vaikuttavuutta muiden tutkijoiden tutkimukseen.

Viittausindikaattoreiksi on valittu normalisoitu viittausindeksi ja top 10 -indeksi. Nämä viittausindikaattorit ovat olleet kansainvälisesti käytössä useita vuosia menetelmällisesti laadukkaissa bibliometrisissä analyysissä. Vastaavia viittausindikaattoreita on käytetty esimerkiksi NordForskin (2014¹) Pohjoismaiden yliopistojen bibliometriikkaan perustuneessa vertailussa, pelkästään bibliometriikkaan perustuvassa yliopistojen Leiden rankingissä² sekä ruotsalaisessa huippututkimuksen kansainvälisessä vertailussa (Karlsson & Persson 2012³). Eri analyysissä aineiston käsittelyssä ja laskentaperusteissa voi olla joitakin eroavaisuuksia, mutta peruseriaatteet ovat samat.

Miksi viittausmäärä normalisoidaan?

Viittauskäytännöt vaihtelevat eri tieteenaloilla: toisaalta se, kuinka moneen aikaisempaan julkaisuun on tapana viitata ja se, kuinka nopeasti ja kuinka pitkän ajan kuluessa julkaisuihin viitataan. Viitetietokannan aineiston päivittyessä viittausindikaattoreiden arvot voivat muuttua. Myös julkaisutyyppi voi vaikuttaa viittauskertymään. Näistä syistä bibliometristen viittausindikaattoreiden laskennassa julkaisujen sama viittausmäärä normalisoidaan tieteenalan (Thomson Reuters subject category), julkaisutyyppin (esim. alkuperäisartikkeli ja katsausartikkeli)

¹ NordForsk (2014). Comparing Research at Nordic Universities using Bibliometric Indicators. Second report, covering the years 2000–2012. NordForsk Policy Paper 2 –2014.

² <http://www.leidenranking.com/>

³ Karlsson Staffan and Persson Olle (2012). The Swedish production of highly cited papers. Vetenskapsrådets lilla rapportserie 5:2012. Julkaisussa Öquist, Gunnar and Benner, Mats (2012). AKADEMIRAPPORT Fostering breakthrough research: a comparative study. Kungl. Vetenskapsakademien.

Saatavilla: http://www.kva.se/globalassets/vetenskap_samhallet/forskningspolitik/2012/akademirapport_breakthrough_research_121209.pdf

sekä julkaisuvuoden mukaan. Esimerkiksi Suomessa tehtyjä julkaisuja verrataan kansainväliseen tasoon samalla tieteenalalla, samassa julkaisutyypissä ja samana julkaisuvuonna.

Miksi ja miten julkaisut ositetaan?

Julkaisut ositetaan maiden, Suomen tutkimusorganisaatioiden ja tieteenalojen kesken. Viittausindikaattorit on skaalattu siten, että maailman keskitaso kullakin tieteenalalla on aina yksi. Yksi julkaisu esiintyy laskennassa vain yhden kerran eivätkä usean organisaation yhteisjulkaisut tule mukaan laskentaan useana kappaleena.

Julkaisujen ositus maiden kesken tarkoittaa sitä, että esimerkiksi suomalais-ruotsalainen julkaisu tuottaa 0,5 julkaisupistettä molemmille maille. Jos tähän julkaisuun on osallistunut tutkijoita kolmesta Suomen yliopistosta, kukin organisaatio saa $1/3 \times 0,5$ julkaisupistettä. Ositus tehdään myös tieteenalojen kesken. Yksi julkaisu voi saada 1–6 tieteenalaa (subject category) sen mukaan, mille tieteenaloille Thomson Reuters on luokitellut julkaisukanavan (tieteellisen lehden).

- Maakohtaisia analyyseja varten jokainen julkaisu ositetaan tasan kaikkien sen edustamien tieteenalojen (= Thomson Reutersin lehdelle määrittelemä subject category) ja sen osoitetiedoissa mainittujen maiden suhteen.
- Jos kustantaja on antanut julkaisulle D kappaletta tieteenaloja ja tekijät tulevat M eri maasta, niin yhden maan osuus yhdellä tieteenalalla on $f = \frac{1}{M \cdot D}$.
- Kotimaisten organisaatioiden vertailua varten Suomen maaosuus on edelleen ositettu tasan kaikkien tunnistettujen organisaatioiden kesken.
- Jos organisaatioita on N kappaletta, niin yhden organisaation osuus yhdellä tieteenalalla on $f = \frac{1}{M \cdot D \cdot N}$.

Mitä normalisoitu viittausindeksi mittaa ja miten se lasketaan?

Normalisoitu viittausindeksi mittaa sitä, kuinka paljon enemmän tai vähemmän viittauksia kuin maailmassa keskimäärin tieteenalan tai tutkimusorganisaation julkaisut tietyllä tieteenalalla ovat tiettynä ajanjaksona keränneet. ”Maailman” eli Thomson Reutersin tietokantoihin indeksoitujen julkaisujen mukaan tieteenalan keskimääräinen taso on 1.

WoS-tieteenala: Thomson Reuters Subject category

Julkaisutyyppi: Article, Letter, Review

Analyysiyksikkö: Suomen tutkimusorganisaatio (esim. yliopisto) tai organisaatioryhmä (esim. yliopistosektori); eri maiden vertailussa yksittäinen maa

Perusluokka: Tietyn julkaisutyyppin julkaisut tietyssä Thomson Reuters ”subject categoryssä” tiettynä vuonna

Normalisoidun viittausindeksin laskennassa julkaisut käsitellään ositettuina. Olkoon I tarkasteltavan analyysiyksikön ositettujen julkaisujen indeksijoukko. Yksi todellinen julkaisu voi siis esiintyä tässä joukossa useampaan kertaan, jos se edustaa useampaa eri WoS-tieteenalaa. Julkaisun $i \in I$ edustamalla alalla julkaisulle lasketaan ensin normalisoitu viittausluku r_i jakamalla sen saama viittausmäärä c_i saman WoS-tieteenalan, julkaisuvuoden ja julkaisutyyppin määrittelemän perusluokan G_i keskimääräisellä viittausmäärällä a_i , joten $r_i = \frac{c_i}{a_i}$. Olkoon edelleen f_i julkaisun ylempänä esitetyllä tavalla laskettu WoS-tieteenalakohtainen osuus. Analyysiyksikön

suhteellinen viittausindeksi ρ saadaan laskemalla osuuksilla painotettu aritmeettinen keskiarvo yksikön kaikkien julkaisujen normalisoiduista viittausluvuista: $\rho = \frac{\sum_{i \in I} f_i * r_i}{\sum_{i \in I} f_i}$.

Mitä top 10 -indeksi mittaa ja miten se lasketaan?

Tieteellisesti vaikuttavimman tutkimuksen osuutta voidaan analysoida tarkastelemalla julkaisuja, jotka kuuluvat tieteenalallaan maailman eniten viittauksia saaneeseen kymmeneen prosenttiin kaikista julkaisuista. Viittausindikaattoriksi valittu top 10 -indeksi kuvaa sitä, kuinka paljon enemmän tai vähemmän kuin maailmassa keskimäärin tieteenalan tai tutkimusorganisaation julkaisuista kuuluu eniten viitattuun kymmeneen prosenttiin tieteenalalla.

Top 10 -indeksi lasketaan samalla periaatteella kuin normalisoitu viittausindeksi, mutta nyt normalisoitu viittausluku r_i korvataan pistemäärällä $p_i > 0$, jos julkaisu kuuluu perusluokan G_i viitatuimpaan kymmenesosaan, ja muulloin arvolla 0. Luvun p_i arvo riippuu luokan G_i koosta ja luokan viitatuimpaan kymmenykseen kuuluvien julkaisujen lukumäärästä, ja se valitaan kussakin luokassa siten, että luokan pistekeskiarvoksi tulee tasan 1,0.

Miten pieni julkaisumäärä vaikuttaa top 10 -indeksiin?

Top 10 -indeksi esitetään vain silloin, kun tieteenalan nelivuotiskauden ositettu julkaisumäärä on vähintään 50. Tämä on yleisesti bibliometrisissä viittausanalyysissä käytetty ositetun julkaisumäärän minimi. Rajasta huolimatta julkaisumääriltään pienillä tieteenaloilla top 10 -indeksin arvot voivat vaihdella paljon tarkastelukausten välillä. Tämä ei kuitenkaan tarkoita, että tieteenalan tutkimuksen taso vaihtelisi muutamassa vuodessa suuresti.

Esimerkki:

- Oletetaan, että Suomen yliopistoissa tehdään tietyllä tieteenalalla 80 julkaisua nelivuotiskaudella.
- Maailman keskitaso eli top 10 -indeksin arvo 1 saavutetaan 8 sellaisella julkaisulla, jotka sijoittuvat tieteenalalla eniten viittauksia keränneiden 10 prosentin joukkoon.
- Jos tällaisia julkaisuja on 7, top 10 -indeksin arvo on 0,88 eli Suomen yliopistoissa tehdyt julkaisut sijoittuvat maailman keskitason alapuolelle tällä tieteenalalla.
- Jos tällaisia julkaisuja on puolestaan 12, top 10 -indeksin arvo on 1,50 eli Suomen yliopistoissa tehdyt julkaisut sijoittuvat selkeästi maailman keskitason yläpuolelle tällä tieteenalalla.
- Siten pienet vaihtelut paljon viitattujen julkaisujen määrässä vaikuttavat tuloksiin paljon.
- Koska top 10 -indeksin laskenta perustuu maiden, organisaatioiden ja tieteenalojen kesken ositettuun julkaisumäärään, todellisuudessa julkaisuja tarvitaan usein enemmän kuin edellä on esitetty.

Lisätietoja:

www.aka.fi/tieteentila > Menetelmät ja luokitukset