

*TEKNISTIETEELLISEN TIEDON
NÄKYVYYDEN MITTAAMINEN:
ESISELVITYS MAHDOLLISUUKSISTA*

*Irma Pasanen
TKK kirjasto*

*Pirjo Sutela
VTT Tietoratkaisut*

RAPORTTI

Huhtikuu 2008

“Lähdeaineiston valossa ja sen sisältämän luokituksen mukaisesti...”

1. Johdanto.....	2
2. Hankkeen tausta	2
2.1. Hankkeen tavoitteet ja osatavoitteet	3
2.2. Hankkeen osallistujat.....	3
2.3. Tarkastellut tietokannat	3
2.3.1. Scopus.....	3
2.3.2. ISI Thomson Scientific.....	7
2.3.3. Hirsch-indeksi eli h-indeksi	12
2.3.4. Julkaisurekisterit TKKJULKAISEE ja VTT:n julkaisurekisteri	13
3. Toimenpiteet.....	14
4. Tulokset.....	16
4.1. TKK	16
4.2. VTT.....	19
4.3. Tietotekniikan alan konferenssit.....	22
5. Johtopäätökset ja julkaisuihin perustuvien arviointien ajankohtaisuudesta	24
5.1. Esiselvityksen tulosten osviitat	24
5.2. ”Measuring Science”- kurssi	26
5.3. Ranking.....	27
6. Esiselvityksen tuloksen arviointi.....	28
Lähteet.....	29
LIITE 1.....	30

1. Johdanto

Teknillisen korkeakoulun TKK:n ja Valtion teknillisen tutkimuskeskuksen VTT:n yhteistyössä tehdyssä hankkeessa vertailtiin Thomson ISI Scientific Web of Science (WoS) ja Scopus tietokantoja. Hankkeessa selvitettiin tietokantojen päällekkäisyyttä ja eroja viittausten määrässä ja laadussa.

Hanke oli luonteeltaan esiselvitys, jonka tarkoituksena oli tuottaa päätöksenteon tueksi tietoa siitä, miten tekniikan alan tutkimukseen liittyvän julkaisutoiminnan erityispiirteet huomioidaan eri tietokantojen pohjalta tehtävissä bibliometrisissa tarkasteluissa. Aineistona esiselvityksessä käytettiin TKK:n ja VTT:n eri aloja edustavien tutkijoiden julkaisutoimintaa koskevia tietoja vuosina 2000 - 2006 sekä heidän vastaavana ajankohtana Thomson Scientific ja Scopus tietokantoihin referoituja julkaisujaan. Lisäksi selvitettiin kirjallisuuden avulla toisaalla jo tehtyjä aiheeseen liittyviä tutkimuksia ja verrattiin saatuja tuloksia niihin.

Viittaustietoja kokoavat myös Google Scholar (<http://scholar.google.com>) ja erilaiset aihealueittain kootut hakukoneet kuten tietotekniikan alan Citeseer (<http://citeseer.ist.psu.edu/>), mutta nämä jätettiin tämän tarkastelun ulkopuolelle. Tällaiset tietokannat ovat vapaasti verkossa kaikkien saatavilla, ja ne referoivat verkossa olevaa materiaalia. Ne eivät siten ole suoraan vertailukelpoisia ISI Thomson Scientific WoS ja Scopus tietokantojen kanssa.

2. Hankkeen tausta

Erilaiset luokittelut tutkimuksen kansainvälisestä paremmuusjärjestyksestä ovat lisääntyneet globaalien kilpailun myötä. 2000-luvulla käynnistyneitä yliopistoja koskevia ranking-listoja ovat mm. THES-QS World University Rankings ja Academic Ranking of World Universities. Ensiksi mainittu tunnetaan Times Higher Education Supplement-luokitteluna ja jälkimmäinen ns. Shanghain listana. Yksi arviointikriteeri näissä arvioinneissa on instituutioiden tuottaman julkaisutoiminta.

Suomessa Suomen Akatemia on laatinut Suomen tieteen tilaa ja tasoa arvioivia katsauksia, joissa jossain vertailtiin Suomen ja muiden OECD-maiden tutkimuksen tuloksellisuutta ja näkyvyyttä. Tuloksellisuudella on tarkoitettu kansainvälisten julkaisujen lukumäärää ja vaikuttavuudella julkaisujen saamia viittauksia. Tekniikan alan tutkimus ei näissä vertailuissa ole saavuttanut hyvää asemaa. Tuloksen perusteella Suomen Akatemian pääjohtaja Raimo Värynen kyseenalaistikin tekniikan tutkimuksen saaman julkisen tuen sen tuottaman tieteellisen panoksen valossa raportin julkistamistilaisuudessa Helsingissä 19.12.2006 (Aamulehti 20.12.2006).

Suomen Akatemian selvityksissä on käytetty National Citation Report NCR ja National Science Indicators NSI tietokantoja (ISI Thomson Scientific). On tunnettua, että Thomson Scientific tietokantojen sisältö painottuu angloamerikkalaiseen lääketieteeseen,

fysiikan ja kemian alan julkaisuihin. ISI Thomson Scientific viittaustietokannoilla on ollut monopoliasema viittaustietojen tuottajana. Vuonna 2004 markkinoille saatiin kilpailija kun Elsevier julkisti Scopus tietokantansa, joka kattaa enemmän kansainvälistä aineistoa kuin ISI Thomson Scientific tietokannat. Lisäksi Scopus sisältää paljon konferenssijulkaisuja.

2.1. Hankkeen tavoitteet ja osatavoitteet

Tekniikan alalla konferenssit muodostavat tieteellisen tiedonvälityksen kannalta hyvin tärkeän foorumin. Laatukongresseissa julkaistaan tutkimusten alkuperäistuloksia ja hyväksytyt täysimittaiset paperit käyvät läpi arviointimenettelyn. Opetusministeriön KOTA-luokitustietoihin onkin toivottu muutosta, jotta tällaiset referoidut konferenssijulkaisut rinnastettaisiin referoituihin journal-artikkeleihin.

Hankkeen tavoitteena on tuottaa päätöksenteon tueksi tietoa siitä, kuinka tekniikan alan tutkimukseen liittyvän julkaisutoiminnan erityispiirteet huomioidaan ISI Thomson Scientific ja Scopus tietokannoissa. Hankkeessa on selvitetty tietokantojen päällekkäisyyttä ja eroja viittausten määrässä ja laadussa. Laatua voidaan tarkastella myös uuden tulokellisuuden mittarin, ns. Hirsch-indeksin avulla, joka on myös laskettavissa kummassakin tietokannassa. Kirjallisuuden avulla selvitetään lisäksi toisaalla jo tehtyjä aiheeseen liittyviä tutkimuksia ja verrataan saatuja tuloksia niihin.

2.2. Hankkeen osallistajat

Teknillisen korkeakoulun TKK:n ja Valtion teknillisen tutkimuskeskuksen VTT:n yhteistyössä tehdyssä hankkeessa on vertailtu ISI Thomson Scientific WoS ja Scopus tietokantoja. Aineistona tutkimuksessa käytetään TKK:n ja VTT:n eri aloja edustavien tutkijoiden julkaisutoimintaa koskevia tietoja vuosina 2000-2006 sekä heidän vastaavana ajankohtana WoS ja Scopus tietokantoihin referoituja julkaisujaan.

2.3. Tarkastellut tietokannat

2.3.1. Scopus

Scopus on monitieteinen, laaja tiivistelmä- ja viittaustietokanta, joka kattaa 15 000 tieteellistä lehteä yli 4 000 kustantajalta. Mukana on myös seuraavanlaisia aineistoja: Open Access-lehtiä (yli 1 000 nimekettä), konferenssijulkaisusarjoja (yli 500 sarjaa), kaupallisia ja ammattijulkaisuja (yli 600, joiden artikkeleita referoidaan valikoidusti) ja kirjasarjoja (yli 125 nimekettä), yhteensä yli 33 miljoonaa viitettä.

Scopus referoi sekä ISSN-numeron sisältäviä vakiintuneita konferenssisarjoja että myös sellaisia konferenssisarjoja, joilla ei ole ISSN-numeroa. Myös Thomson Scientific

tietokannat referoivat konferenssisarjoja, mutta ainoastaan sellaisia, joilla on ISSN-numero.

Viittaukset ovat mukana tietokannassa vuodesta 1996 lähtien. Koko aineisto takautuu kuitenkin huomattavasti kauemmaksi, vaihdellen tiedontuottajittain.

Scopusin maantieteellinen kattavuus on maailmanlaajuinen ja siinä on vahva paino muulla kuin pohjoisamerikkalaisella kirjallisuudella: yli puolet sisällöstä on peräisin Euroopasta, Latinalaisesta Amerikasta ja Kaakkois-Aasiasta ja noin 35 % Pohjois-Amerikasta.

Maantieteellisestä kattavuudesta kertoo myös se, että De Moya-Anegonin [2007] selvityksen mukaan 15 % Scopuksessa esiintyvistä julkaisuista on muun kuin englanninkielisiä. Selvityksessä verrattiin Scopusin kattavuutta Ulrichs-lehtiluetteloon. Muun kuin englanninkielisiltä julkaisuilta edellytetään kuitenkin englanninkielistä abstraktia.

Scopusin kattavuus aiheryhmittäin on esitetty taulukossa 1.

Scopusin aiheryhmät ja niiden kattavuus	
<p>Life Sciences 3 400 nimekettä 20 % kattavuudesta</p> <ul style="list-style-type: none"> • Agricultural and biological sciences • Biochemistry, Genetics and Molecular Biology • Immunology and Microbiology • Neuroscience • Pharmacology, Toxicology and Pharmaceutics 	<p>Social Sciences 2 850 nimekettä 17 % kattavuudesta</p> <ul style="list-style-type: none"> • Arts and humanities • Business, Management and accounting • Decision Sciences • Economics, Econometrics and Finance • Psychology • Social Sciences
<p>Physical sciences 5 500 nimekettä 32 % kattavuudesta</p> <ul style="list-style-type: none"> • Chemical engineering • Chemistry • Computer Science • Earth and Planetary science • Energy • Engineering • Environmental Science • Materials Science • Mathematics • Physics and Astronomy 	<p>Health Sciences 5 300 nimekettä 31 % kattavuudesta</p> <ul style="list-style-type: none"> • Medicine • Nursing • Veterinary • Dentistry • Health professionals

Lähde: <http://www.info.scopus.com/>

Taulukko 1. Scopusin aiheryhmät ja niiden kattavuus.

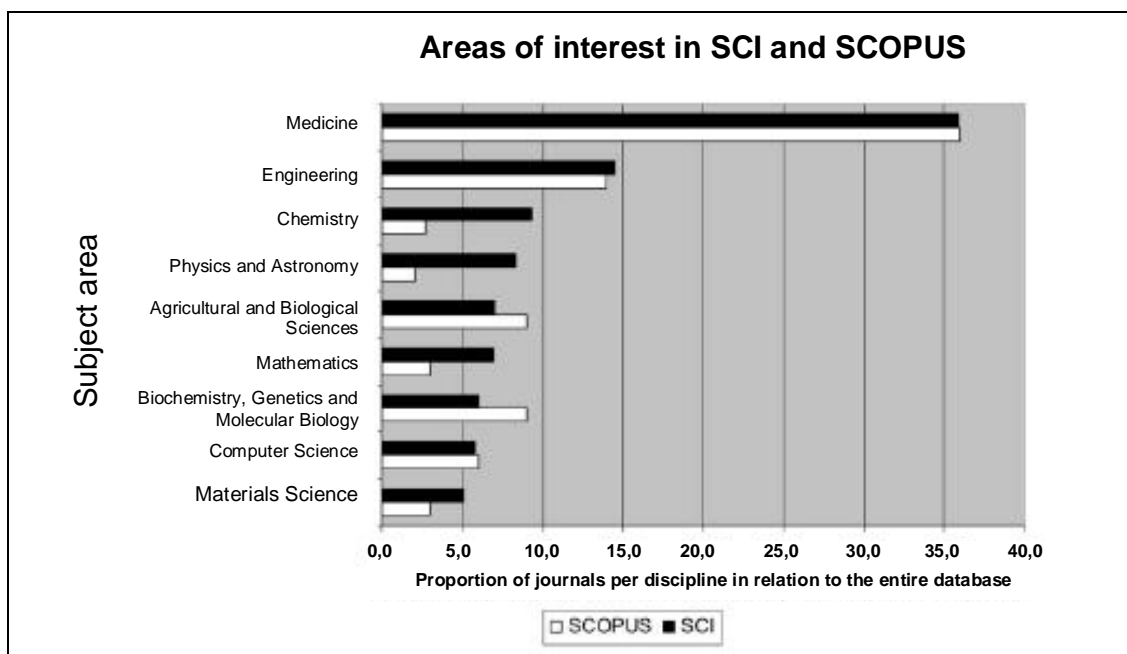
Scopusin aiheluokkien kattavuudesta on tehty myös selvityksiä. Dessin [2006] testihakuihin perustuvan selvityksen mukaan kattavuus on seuraava:

- Health 14.3 miljoonaa dokumenttia
- Life Sciences 11.0
- Engineering 8.5
- Agricultural & Biological Sciences 3.6
- Earth & Environmental Sciences 1.9
- Chemistry 1.3
- Physics 0.59
- Social Science 0.29
- Mathematics 0.26
- Psychology 0.23
- Economics, Business, Management 0.22

Selvityksen mukaan Scopus on selvästi vahvin terveystieteissä, biotieteissä ja tekniikassa. Sen sijaan kattavuus on selvästi heikompi maatalouden, kemian, fysiikan, matematiikan, psykologian ja yhteiskuntatieteiden aloilla.

Scopusin vahvuus eri tieteenaloilla tulee suureksi osaksi sen pohjatietona olevista alakohtaisista pääasiassa Elsevierin tuottamista tietokannoista. Näitä ovat Embase: lääketiede, Medline via Embase (tuottaja on National Library of Medicine, USA): lääketiede, Compendex: laaja teknisten alojen tietokanta, World Textile Index: tekstiilit, Geobase: geotieteet, Biobase: biotieteet ja Fluidex: nesteet [Bosman et al. 2006]. Scopusin selvin vahvuus myös näiden taustalla olevien tietokantojen valossa on lääketieteen, terveyden, teknisten tieteiden ja biotieteiden aloilla.

Ball ja Dunger [2006] ovat vertailleet Thomson Scientific ja Scopus –tietokantojen kattavuutta eri tieteenaloilla. Heidän mukaansa jokainen tietokanta kehittää oman fokuksensa koskien tieteenalaa, kieltä ja maantieteellistä kattavuutta. Tietokantojen keskittymistä eri tieteenaloille on selvityksessä vertailtu eri aloille kuuluvien lehtien osuuden avulla (kuva 1). Vertailuista käy ilmi, että molempien tietokantojen suurin painopiste on lääketieteessä. Suurimmat erot löytyvät kemiasta, fysiikasta ja astronomiasta sekä matematiikasta.



Kuva 1. Scopus ja Thomson Scientific –tietokantojen kattamien aihepiirien vertailu [Ball ja Dunger 2006].

Scopusen referoimat **keskeiset dokumenttityypit** on esitetty taulukossa 2.

Scopusen keskeiset dokumenttityypit	
Dokumenttityyppi	Kuvaus
Article	Original research or opinion, also includes conference papers
Book	Whole monograph or a chapter of a book
Conference report	Additional item summarizing all papers from a conference
Report	A non-serial publication giving a detailed account of information or statements, often including opinion and findings, of an individual or group on a particular topic
Review	Significant review of original research, also includes conference papers
Short survey	Short or mini-review of original research

Lähde: <http://www.info.scopus.com/>

Taulukko 2. Scopusen keskeiset dokumenttityypit.

Scopus julkisti loppuvuonna 2007 laajoja bibliometrisia mittauksia varten tarkoitetun uuden tuotteen, Scopus Custom_data, joka tarjoaa käyttäjälle räätälöityjä otteita Scopus -tietokannasta XML-muodossa. Tuotteeseen kuuluu myös analysointiohjelmisto.

2.3.2. ISI Thomson Scientific

2.3.2.1 Web of Science (WoS) eli Science Citation tietokannat

ISI Thomson Scientificin tieteellisen julkaisemisen tietokantaperheen ytimen muodostavat kolme aihealueittain koottua tietokantaa:

1. Science Citation Index (SCI). Tietokantaan referoidaan noin 5900 lääketieteen, luonnontieteiden ja tekniikan alan aikakauslehteä.
2. Social Science Citation Index (SSCI). Tietokantaan referoidaan noin 1700 talous- ja yhteiskuntatieteiden alan aikakauslehteä.
3. Arts and Humanities Citation Index (AHCI). Tietokantaan referoidaan noin 1130 humanististen ja taidealojen aikakauslehteä.

Web of Science, WoS, on yllämainittuja tietokantoja koskeva verkkopalvelu ja se on osa laajempaa Web of Knowledge – kokonaisuutta.

Tietokantoihin referoidut lehdet ovat lähinnä angloamerikkalaisia. Räätelöityjä poimintoja näistä tietokannoista ovat mm. maakohtainen National Citation Report NCR, National Science Indicators NSI ja Essential Science Indicators ESI.

ISI Thomson Scientificin tietokantojen referoimat **keskeiset dokumenttityypit** on esitetty taulukossa 3.

WoS keskeiset dokumenttityypit	
Dokumenttityyppi	Kuvaus
Article	Ei löydy tarkentavaa määritystä
Meeting abstract	Meeting Abstracts are covered for the top 900 journals (ranked by impact factor) in Science Citation Index Expanded. All meeting abstracts appearing journals in the Social Sciences Citation Index and the Arts & Humanities Citation Index are indexed.
Review	Articles are coded as reviews if they are either published in a review journal or include 100 or more cited references along with some indicator that the article is a review, such as the word "Review" or "Overview" in the table of contents or in the article itself.
Editorial material	Ei löydy tarkentavaa määritystä
Letter	Ei löydy tarkentavaa määritystä
Book review	In Science Citation Index Expanded, only those book reviews that appear in Science, Nature, and The Scientist are covered. Book reviews are fully covered in the Social Sciences Citation Index and Arts and Humanities Citation Index.
Correction	Ei löydy tarkentavaa määritystä

Lähde: Web of Science 8.0 Workshop 2007

<http://scientific.thomson.com/media/scpdf/ws-wos-8-0-0807.pdf>

Taulukko 3. WoS tietokantojen keskeiset dokumenttityypit.

Science Citation Index tietokanta keskittyy erityisesti fysiikan, kemian ja lääketieteen alojen julkaisuihin. Tarkasteltaessa esimerkiksi viittauksia ISIn referoimiin aikakauslehtiin on todettu, että kemian ja fysiikan aloilla yli 80 % kaikista viittauksista saadaan ISIn lähdeaineistona olevista aikakauslehdistä, ja näin tapahtuu myös biotekniikassa (taulukko 4). Sen sijaan matematiikassa, insinööritieteissä ja taloustieteissä on tämä kattavuus 40 - 60 %. Tietyillä aloilla, kuten rakennus- ja liikennetekniikassa, robotiikassa ja tietojenkäsittelyssä, jää kattavuus vielä tätäkin vähäisemmäksi.

ISI THOMSON SCIENTIFIC TIEKANTOJEN (WoS) KATTAVUUS TEKNIIKAN ERI ALOILLA	Aikakauslehtien merkitys alan tiedonvälityksessä: viittaukset aikakauslehdissä julkaistuihin artikkeleihin % osuutena kaikista viittauksista	ISI:n kattavuus aikakauslehtiaineistosta: viittaukset ISI lähdeaineistona oleviin aikakauslehtiin % osuutena kaikista aikakauslehtiin tehdyistä viittauksista	Yleinen kattavuus: viittaukset ISI lähdeaineistona oleviin aikakauslehtiin % osuutena kaikista viittauksista
INSINÖÖRITIETEET			
Tietotekniikka, tekoäly	53	77	41
Tietojenkäsittelyoppi	45	70	31
Rakennustekniikka	51	71	36
Sähkö- ja tietoliikennetekniikka	65	83	54
Konetekniikka	67	76	51
Ydintekniikka	69	85	58
Robotiikka	49	67	33
Liikennetekniikka	37	57	21
SOVELLETTU KEMIA JA FYSIIKKA			
Akustiikka	77	84	65
Kemian tekniikka	77	87	66
Sovellettu kemia	83	85	71
Instrumentointi	72	86	63
Materiaalitiede	83	89	74
Mekaniikka	75	84	63
Optiikka	86	91	78
Sovellettu fysiikka	87	93	81
KEMIA			
Kemia	91	93	84
Analyttinen kemia	89	92	82
Epäorgaaninen ja ydinkemia	89	94	84
Orgaaninen kemia	93	95	88
Fysikaalinen kemia	89	95	85
Sähkökemia	87	91	79
Polymeeritiede	87	91	79
Biotekniikka	90	93	84
FYSIIKKA			
Fysiikka yl.	87	93	81
Kiinteän olomuodon fysiikka	91	96	87
Plasmafysiikka	86	93	80
Laskennallinen fysiikka	82	92	76
Atomi- ja molekyyli-fysiikka	90	96	87
Ydinfysiikka	89	93	83
Hiukkasfysiikka	90	92	83

MATEMATIIKKA			
Matematiikka	72	68	49
Sovellettu matematiikka	70	77	54
TALOUSTIETEET			
Liiketalous	64	78	50
Rahoitus	66	83	55
Taloustiede	56	83	47
Johtaminen	59	76	45

Lähde: Citation analysis in research evaluation / Henk F. Moed. Springer 2005 s. 129-130.

Taulukko 4. WoS tietokantojen kattavuus lähinnä tekniikan ja sen lähialoilla.

2.3.2.2 ISI Proceedings

Konferenssiesityksiin keskittynyt ISI Proceedings on jaettu kahteen eri tietokantaan:

1. Science & Technology edition, joka sisältää tietoja lääketieteen, luonnontieteiden ja tekniikan alan konferenssiesityksistä.
2. Social Sciences & Humanities edition sisältää tietoja humanististen alojen sekä talous- ja yhteiskuntatieteiden alan konferenssiesityksistä.

Tiedot on jaoteltu 250 eri luokkaan, kuten esim. insinööritieteet, fysiikka, kemia, biologia, psykiatria ja taloustieteet. ISI Proceedings tietokannoissa on tietoja vuodesta 1990 alkaen ja vuosittain tietokantaa kartutetaan noin 385 000:lla konferenssiesityksellä.

Viittaustietoja konferenssiesityksistä artikkeleihin tai konferenssiesitysten välille ei koota ISI proceedings tietokantaan.

2.3.2.3 Essential Science Indicators ESI

Essential Science Indicators (ESI) sisältää valmiiksi analysoidussa muodossa rankinglistoja tutkijoista, yliopistoista ja tutkimuslaitoksista, maakohtaisen analyysin sekä julkaisukohtaisen analyysin. Ainoastaan ISI Thomson Scientific tietokantoihin indeksoidut lehdet huomioidaan ESI-tietokannan rankinglaskelmissa: sekä artikkelin että siihen tehdyt viittaukset on oltava tietokantaan indeksoiduissa lehdissä. Tällaisia aikakauslehtiä ESI-tietokannassa on kaikkiaan 5 051 kappaletta. Aihealueittaiset rankingluokat perustuvat kuitenkin vielä huomattavasti pienempiin otoksiin, kuten taulukko 5 osoittaa.

<i>ESI aihealue</i>	<i>Lehdet, joiden artikkelit huomioidaan (lukumäärä)</i>
<i>Agricultural Sciences</i>	<i>121</i>
<i>Biology & Biochemistry</i>	<i>242</i>
<i>Chemistry</i>	<i>305</i>
<i>Clinical Medicine</i>	<i>891</i>
<i>Computer Science</i>	<i>174</i>

<i>Economics & Business</i>	199
<i>Engineering</i>	452
<i>Environment/Ecology</i>	149
<i>Geosciences</i>	185
<i>Immunology</i>	57
<i>Materials Sciences</i>	170
<i>Mathematics</i>	181
<i>Microbiology</i>	64
<i>Molecular Biology & Genetics</i>	132
<i>Multidisciplinary</i>	19
<i>Neuroscience & Behavior</i>	131
<i>Pharmacology & Toxicology</i>	94
<i>Physics</i>	175
<i>Plant & Animal Science</i>	375
<i>Psychology/Psychiatry</i>	276
<i>Social Sciences, general</i>	620
<i>Space Science</i>	39

Taulukko 5. Essential Science Indicators -tietokannan aihealueet ja niiden sisältämien aikakauslehtien lukumäärät.

Ranking-listalle pääsy edellyttää siten julkaisemista tietyssä lehdessä, ja että ao. artikkeliin on myös viitattu lehdessä, jota ISI Thomson Scientific indeksoi. Lisäksi ranking-listalle pääsy edellyttää saatuihin viittauksiin perustuvan aihealuekohtaisen kynnysarvon ylittämistä.

Kaikki artikkelin kirjoittajat saavat yhtäläisen meriitin sekä tekijän että viittausten osalta. Viittaukset poimitaan ISI Thomson Scientific-tietokantoihin indeksoiduista lehdistä, ja hyväksytyjä artikkelityyppejä ovat tieteelliset artikkelit, katsaukset, konferenssiesitykset ja tutkimusta koskevat tiedotteet. Tietojen poiminnan ajanjakso on 10 vuotta.

Lehtien perusteella tapahtuva aihealueluokittelu ryhmittelee tieteenalat 22 eri luokkaan, ja yksi lehti voi kerrallaan olla vain yhdessä luokassa. Monialaisten lehtien luokka sisältää lehdet kuten Science and Nature, ja tämän luokan sisältämien lehtien artikkelit jaetaan edelleen muihin luokkiin artikkelikohtaisesti. Tämä tulee ottaa huomioon kun arvioidaan yksittäisen tutkijan, laitoksen tai maan sijoittumista ranking-listalla.

2.3.2.4 Journal Citation Reports JCR

Journal Citation Report JCR pohjaa 1970-luvulta alkaen ISI-tietokantojen yhteydessä kerättyyn tietoon julkaisuista ja niiden saamista viittauksista. JCR ilmestyy vuosittain kahdessa osassa,

- Science-osa sisältää tietoja noin 5 900 lääketieteen, tekniikan ja luonnontieteiden lehdestä
- Social Sciences-osa sisältää tietoja noin 1 700 talous- ja yhteiskuntatieteiden lehdestä

JCR lehdet ovat lähinnä angloamerikkalaisia ja niille on tietokannassa laskettu vaikuttavuuskerroin, (IF, Impact Factor). Suurimmat vaikuttavuuskertoimet ovat lääketieteen lehdillä ja suurilla tieteen yleislehdillä Science and Nature.

Tieteellisten aikakauslehtien keskinäistä paremmuutta arvioidaan usein tämän vaikuttavuuskertoimien avulla. Mitä suurempi vaikuttavuuskerroin, sitä enemmän lehdessä ilmestyneet artikkelit ovat saaneet viittauksia. Kun vain kolmannes kaikista tieteellisistä julkaisuista tulee koskaan saamaan edes yhden viittauksen, pidetään huippulehdessä ilmestyneen artikkelin mahdollisuuksia tulla luetuksi ja myöhemmin viitatuksi suurempana kuin vähemmän tunnetun lehden artikkelin mahdollisuuksia.

Vaikuttavuuskertoimia voi siis käyttää esimerkiksi valittaessa tutkimuksen julkistamisfoorumia. Usein myös rahoitushakemuksissa mainitaan, mikäli aikaisempia tutkimukseen liittyviä julkaisuja on ilmestynyt korkean vaikuttavuuskertoimen lehdissä.

Vaikuttavuuskerroin lasketaan jakamalla

lehdessä Y vuonna X ilmestyneiden artikkelien saamien viittausten lukumäärä

kahden kyseistä vuotta aikaisemman vuoden aikana (X-2) lehdessä Y ilmestyneiden artikkelien lukumäärällä.

ISIn vaikuttavuuskertoimia on kritisoitu seuraavista syistä:

- Mukaan laskettavien artikkelien käsite epäselvä: ovatko pääkirjoitukset, kirjeet toimitukselle, arvostelut artikkeleja
- Tieteenalakohtaisia eroja julkaisukäytännöissä ei huomioida
- Viittausten kertymäaika (2 vuotta) on liian lyhyt
- On voitu osoittaa selviä laskuvirheitä vaikuttavuuskertoimen laskemisessa.

2.3.2.5 Journal Use Reports (JUR)

Journal Use Reports JUR on tuore, vuonna 2006 käynnistetty tiedon analysointia varten kehitetty tietokanta, jonka kohderyhmänä ovat tutkimuslaitosten hallinto-organisaatiot ja kirjastot. JUR yhdistää organisaation aikakauslehtien käyttötilastotietoja JCR-tietokannan tietoihin em. lehtien vaikuttavuuskertoimista sekä organisaation tutkijoiden julkaisutietoihin ja saatuihin viittauksiin.

2.3.2.6 ISI Highly Cited Researchers

Highly cited researchers on vapaasti verkossa oleva sivusto tutkijoista, joihin on eniten viitattu vuosien 1981-2005 aikana. Aineisto perustuu Science Citation Index, Social

Science Citation Index ja Arts and Humanities Citation Index tietokantoihin indeksoituihin artikkeleihin ja niiden saamiin viittauksiin. Aineisto on jaettu 21 eri tieteenalaluokkaan, ja kullekin kategorialle lasketaan 250 eniten viittauksia saanutta tutkijaa. Artikkeliaineiston jako luokkiin perustuu siihen, että jokainen tieteellinen lehti on sijoitettu pääsääntöisesti yhteen luokkaan, ja siten kaikki em. lehdessä julkaistut artikkelit edustavat tätä aihealuetta. ISI Thomson Scientificin mukaan on kuitenkin mahdollista, että tietyissä tapauksissa aikakauslehti voi olla sijoitettu myös useampaan luokkaan. Luokittelun tarkoituksena on huomioida eri alojen erilaiset viittauskäytännöt. Korkeimmat viittausluvut ovat fysiikassa, lääketieteessä ja kemiassa, alhaisimmat tietotekniikan alalla. Laskentatapa suosii tutkijaa, jonka julkaisutoiminta keskittyy samaan aihealuetta sijoitettuihin aikakauslehtiin. Mikäli tutkijan julkaisutoiminta hajaantuu useaan luokkaan, lasketaan saadut viittaukset kussakin kategoriassa erikseen, jolloin tutkijan on hyvin vaikea saavuttaa vaadittua kynnyksiarvoa missään aihealuetta. <http://www.isihighlycited.com>

2.3.3. Hirsch-indeksi eli h-indeksi

Hirsch-indeksi kertoo helpolla tavalla tutkijan tuottavuutta ja vaikuttavuutta kuvaavan luvun tietyltä ajanjaksolta. Indeksini esitteli Jorge E. Hirsch vuonna 2005 [Hirsch 2005]. Luvun saa suoraan sekä Web of Science WoS että Scopus-tietokannasta. Indeksini käyttö on yleistynyt viime vuosina.

Tutkijan Hirsch-indeksi on se luku, jossa viittausten määrä kohtaa julkaisun järjestysnumeron, kun julkaisut asetetaan viittausten määrän mukaiseen järjestykseen. [*A scientist has index h if h of his N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each*].

Hirsch-indeksi auttaa vertaamaan saman alan tieteen tekijöitä keskenään. Hirsch-indeksi on käyttökelpoinen vain sellaisten tutkijoiden kohdalla, jotka ovat jo julkaisseet paljon.

2.3.4. Julkaisurekisterit TKKJULKAISEE ja VTT:n julkaisurekisteri

2.3.4.1 TKKJULKAISEE

TKKJULKAISEE on Teknillisen korkeakoulun julkaisurekisteri. Tietokanta sisältää tiedot TKK:n henkilöstön julkaisutoiminnasta alkaen vuodesta 1992.

Rekisterin tarkoituksena on palvella niin korkeakoulua kuin sen ulkopuolisia tahoja TKK:n tutkimus-, julkaisu- ja muusta tieteellisestä toiminnasta tiedottamisessa, sekä korkeakoulun hallintoa tietojen tilastoinnissa ja toiminnan suunnittelussa.

TKKJULKAISEE on julkinen tietokanta ja vapaasti kenen tahansa käytettävissä www-selaimen avulla. Tietokantaa päivitetään jatkuvasti ja vuoden 2007 lopussa se sisälsi lähes 50 000 viitettä.

Teknillisen korkeakoulun julkaisuprofiili 2000-luvulla (taulukko 6) osoittaa, että artikkelien ohella konferenssijulkaisut ovat hyvin tärkeä julkaisufoorumi tekniikan alan tutkimustulosten julkistamisessa.

TKK julkaisut 2000-2007	TKKJULKAISEE 2000-2007			
	Kaikki julkaisut n=27115		Kansainväliset julkaisut n=18054	
Lehtiartikkelit	9295	34 %	7069	39 %
Artikkelit kokoomateoksissa	1758	6 %	677	4 %
Esitelmät konferenssijulkaisuissa	11769	43 %	9623	53 %
Monografiat	705	3 %	154	1 %
Raportit	3068	11 %	300	2 %
Patentit	174	1 %	110	1 %
Ohjelmistot	185	1 %	59	0 %
Muut e-julkaisut	161	1 %	62	0 %

Taulukko 6. Teknillisen korkeakoulun julkaisuprofiili 2000-luvulla

2.3.4.2 VTT:n julkaisurekisteri

VTT:n julkaisurekisteri, julkinen www-tietokanta, sisältää viitetiedot VTT:n julkaisutoiminnasta vuodesta 1984 lähtien. VTT:n omien julkaisusarjojen julkaisut ovat mukana vuodesta 1943. Viitteitä on yli 50 000. Rekisterin tehtävänä on palvella VTT:n tieteellisestä toiminnasta tiedottamisessa ja myös tuottaa VTT:n raportoinnissa ja tuloksellisuuden mittauksessa tarvittavaa tilastoaineistoa.

VTT:n julkaisurekisteriin tallennetaan tiedot julkaisuista julkaisutyypeittäin. Taulukossa 7. on VTT:n julkaisutuotanto vuosilta 2000 – 2007 esitettynä julkaisutyypeittäin, koko julkaisutuotanto ja kansainväliset tieteelliset julkaisut myös omana ryhmänään.

Tieteellisistä julkaisuista noin puolet on tieteellisten konferenssien esitelmiä, eli 48 %.

Pienempi osuus on tieteellisten lehtien artikkeleita, 40 %, ja lisäksi VTT:n tutkijat julkaisevat tieteellisissä kokoomateoksissa ja laativat tieteellisiä erillisjulkaisuja, 12 %.

VTT:n julkaisut 2000 - 2007	VTT:n julkaisurekisteri 2000 – 2007			
Julkaisutyyppi	Kaikki julkaisut		Kansainväliset tieteelliset julkaisut	
Lehtiartikkelit	5 793	35 %	2 823	40 %
Artikkelit kokoomateoksissa	535	3 %	535	7 %
Esitelmät konferenssijulkaisuissa	7 540	46 %	3 354	48 %
Monografiat	1 296	8 %	405	5 %
Raportit	814	5 %		
Patentit	303	2 %		
Muut e-julkaisut	172	1 %		
Yhteensä	16 553	100 %	7 117	100 %

Taulukko 7. VTT:n julkaisuprofiili 2000-luvulla

3. Toimenpiteet

Esiselvityksessä tarkasteltiin TKK:n ja VTT:n eri aloja edustavien tutkijoiden julkaisutoiminnan näkyvyyttä eli julkaisujen määrää ja vaikuttavuutta eli julkaisujen saamia viittausmääriä ISI Thomson Scientific Web of Science (WoS) ja Scopus - tietokannoissa. Tavoitteena oli saada käsitys siitä, onko näiden kahden tietokannan välillä selviä eroja. Lisäksi haluttiin nähdä, minkälaisen kuvan kahdesta eri tietokannasta otettu Hirsch-indeksi antaa tutkijoiden tuottavuudesta ja vaikuttavuudesta

Lähtökohtana esiselvityksessä käytettiin TKK:n ja VTT:n eri aloja edustavien tutkijoiden julkaisutietoja vuosilta 2000-2006 kummankin organisaation oman julkaisurekisterin mukaan ja sen jälkeen heidän vastaavana ajankohtana WoS - ja Scopus -tietokantoihin referoituja julkaisujaan. Molemmissa tietokannoissa tehtiin samanlaiset henkilöhaut helmi-maaliskuun aikana 2007.

Tarkastelun kohteeksi valittiin henkilöitä eikä organisaatioita, koska se mahdollisti tarkan haun ja sen mukanaan tuoman vertailtavuuden. Jos sen sijaan organisaatio tai sen osa olisi valittu tarkasteluun, olisi siihen helposti liittynyt organisaatioiden nimien monenlaisten kirjoitusasujen aiheuttamia kattavuusongelmia.

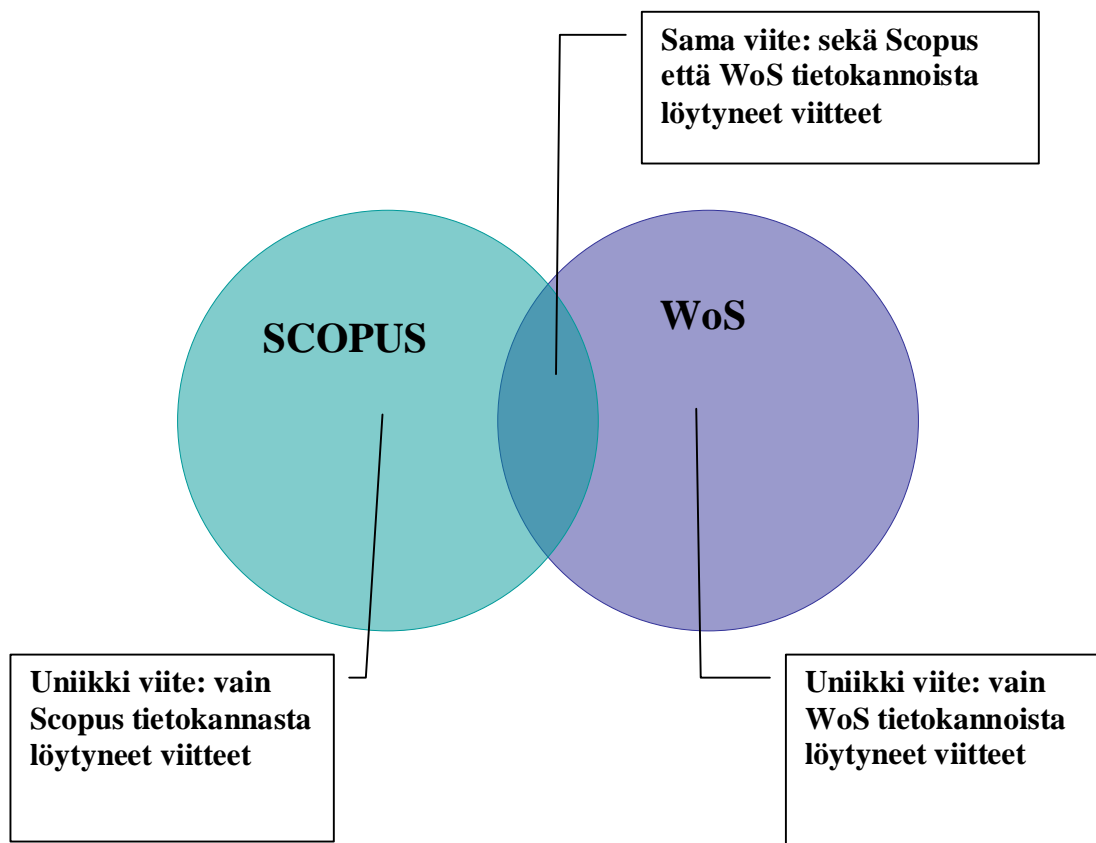
Kuvassa 2 on havainnollistettu tietokantojen päällekkäisyyttä. Tässä esiselvityksessä on käytetty seuraavia käsitteitä päällekkäisyyden kuvaamiseen:

Viite: tietokannasta löytynyt julkaisua koskeva viite

Uniikki viite: viite, joka löytyy vain toisesta tietokannasta

Sama viite: viite, joka löytyy kummastakin tietokannasta ja osoittaa siten päällekkäisyyttä

Yhteensä viitteet: tietokannoista löydetty viitteet yhteensä, samat päällekkäiset viitteet huomioitu vain kertaalleen.



Kuva 2. Tietokantojen unioni, tietojen päällekkäisyys ja käytetyt termit.

4. Tulokset

4.1. TKK

Tarkasteluun valittiin 20 TKK:n eri aloja edustavan alallaan suhteellisen paljon julkaisseen varttuneemman tutkijan julkaisutuotanto. Tutkijoista osa oli alansa huippututkijoita. Aineistosta erotettiin edelleen kansainväliseksi julkaisuksi ilmoitetut julkaisut vuosina 2000-2006. Valittu otos noudatti pääpiirteissään TKK:n yleistä julkaisuprofiilia (taulukko 8), painottuen kuitenkin enemmän lehtiartikkeleihin.

TKK 20 tutkijan kansainvälinen julkaisutuotanto 2000-2006	TKKJULKAISEE 2000-2006 n= 1468 julkaisua	
Lehtiartikkelit	699	48 %
Artikkelit kokoomateoksissa	58	4 %
Esitelmät konferenssijulkaisuissa	676	46 %
Monografiat	17	1 %
Raportit	6	0,5 %
Patentit	10	1 %
Ohjelmistot	1	0 %
Muut e-julkaisut	1	0 %

Taulukko 8. TKK:n 20 tutkijan kansainvälinen julkaisutuotanto 2000-2006 julkaisutyypeittäin.

Kaikkiaan otokseen kuului TKK:n julkaisurekisteriin kansainvälisiksi julkaisuiksi ilmoitetut 1468 julkaisua (taulukko 8). Pääosan muodostivat artikkelit ja esitykset konferenssijulkaisuissa, ne muodostivat 98 % kaikista otoksen julkaisuista. Sekä SCOPUS ja WoS tietokannoissa löytyi viitetiedot hieman yli 700 julkaisusta, jotka eivät kuitenkaan olleet kaikki samoja julkaisuja. Pääallekkäisiä julkaisuja, jotka löytyivät kummastakin tietokannasta, oli 71 % (587 kpl). Uniikkeja viitteitä oli 29 % koko aineistosta, ja se jakaantui tasan Scopus (121 kpl) että WoS (122 kpl) tietokantojen kesken. Erilaisia TKK:n julkaisuja löytyi siten näistä tietokannoista yhteensä 830 kpl (taulukko 9). Tämä vastaa 57 %:a otoksen 1468 julkaisusta.

TKK Julkaisujen määrät ja osuudet: Samat, uniikit ja eri julkaisut Scopus ja WoS tietokannoissa				
Viitteitä	Scopus viitteet	WoS viitteet	Viitteet yhteensä ja osuus koko aineistosta	
Samoja viitteitä	587	587	587	71 %
Uniikkeja viitteitä	121	122	243	29 %
Yhteensä eri viitteitä	708	709	830	100 %

Taulukko 9. Samat, uniikit ja eri TKK:n julkaisut Scopus ja WoS tietokannoissa

Molemmissa tietokannoissa löytyneitä samoja julkaisuja koskeva lähempi tarkastelu on esitetty taulukossa 10. Taulukosta käy ilmi, että kaikkien julkaisujen osalta vaikuttavuus (saadut viittaukset) näyttäisi olevan samankaltainen Scopus ja WoS tietokannoissa. Merkillepantavaa on, että kun tarkastellaan samojen julkaisujen saamia viittauksia, kasvaa ero huomattavasti. Scopusin laaja-alaisempi pohja tuottaa enemmän vaikuttavuutta sillä julkaisut saivat merkittävästi enemmän viittauksia kuin vastaavat julkaisut saivat WoS tietokannoissa. Ero saatujen viittausten määrässä tietokantojen välillä on tilastollisesti melkein merkitsevä ($p=0,041076$). Samansuuntainen havainto tietokantojen välillä (saadut viittaukset tietokannoissa) on tehty myös aikaisemmassa tutkimuksessa informaatiotutkimuksen alalla [Meho and Yang 2007].

TKK Julkaisujen näkyvyys ja vaikuttavuus Scopus ja WoS tietokannoissa		
	Scopus	WoS
Yleinen näkyvyys (löytyvyys tietokannoista)	708	709
Vaikuttavuus (saadut viittaukset tietokannoissa)	4821	4880
Samojen viitteiden saamat viittaukset tietokannoissa		
Samat viitteet	587	587
Samojen viitteiden saamat viittaukset tietokannassa	4584	4037

Taulukko 10. Julkaisujen näkyvyys ja vaikuttavuus Scopusissa ja WoS tietokannoissa sekä samojen viitteiden saamat viittaukset.

TKK Julkaisut aiheoryhmittäin: Samat julkaisut Scopus ja Web of Science tietokannoissa					
Aiheryhmä (Scopus-luokitus)	Scopus-viitteet	WoS-viitteet	Yhteensä	Samoja viitteitä	Samojen viitteiden osuus %
Chemical Engineering	142	129	158	113	72
Computer Science	119	127	154	92	60
Energy	20	16	20	16	80
Engineering	95	66	103	58	56
Environmental Science	6	4	6	4	67
Materials Science	89	109	114	84	74
Mathematics	14	13	14	13	93
Neuroscience	71	85	90	66	73
Physics and astronomy	146	157	164	139	85
Social Sciences (incl. Arts and Humanities)	6	3	7	2	29
Yht.	708	709	830	587	71 %

Taulukko 11. Samojen julkaisujen osuus aihealoittain Scopus ja WoS tietokannoissa.

Taulukossa 11 on esitetty aihealueittainen tarkastelu samojen julkaisujen osalta. Tietojenkäsittelyn ja insinööritieteiden osalta tietokantojen päällekkäisyys on noin 60 %, mikä on selvästi alle keskiarvon, 71%. Yhteiskuntatieteiden ja taidealan sekä ympäristötieteiden osalta otos on liian pieni johtopäätösten tekoon.

TKK:n tutkijoiden saamat Hirsch-indeksit Scopus ja WoS tietokannoissa on esitetty taulukossa 12. Erot ovat pieniä, mutta yleisesti Scopusesta laskettu Hirsch-indeksi on hieman korkeampi kuin WoS tietojen avulla laskettu. Poikkeuksena on kaksi tietojenkäsittelyopin alan tutkijaa, joiden Hirsch-indeksi on korkeampi WoS tietojen avulla laskettuna.

TKK Hirsch-indeksit Scopus- ja WoS-tietokannoissa			
TKK Tutkijan julkaisujen Scopus-aihe ryhmä		Scopus-Hirsch-indeksi	WoS- Hirsch-indeksi
Chemical Engineering	Tutkija 1	8	6
	Tutkija 2	11	11
Computer Science	Tutkija 1	4	6
	Tutkija 2	8	9
	Tutkija 3	9	8
Energy	Tutkija 1	4	3
Engineering	Tutkija 1	3	1
	Tutkija 2	2	1
	Tutkija 3	6	5
	Tutkija 4	2	1
Environmental Science	Tutkija 1	3	3
Materials Science	Tutkija 1	5	5
	Tutkija 2	13	12
Mathematics	Tutkija 1	5	4
Neuroscience	Tutkija 1	19	19
Physics and Astronomy	Tutkija 1	21	20
Social Sciences	Tutkija 1	4	1

Taulukko 12 TKK Hirsch-indeksit Scopus- ja WoS-tietokannoissa

4.2. VTT

Tarkasteluun valittiin 20 VTT:n eri tutkimusaloja edustavaa, VTT:n julkaisurekisterin mukaan suhteellisen paljon julkaisevaa tukijaa ja heidän tieteellinen julkaisutuotantonsa.

Tieteellisiä julkaisuja vuosilta 2000 – 2006 oli VTT:n julkaisurekisterissä selvitykseen valituilla tutkijoilla yhteensä 910 kappaletta. Niistä tieteellisten lehtien artikkeleita oli 451 kappaletta (50 %), konferenssiesitelmiä oli 439 kappaletta (48 %) ja tieteellisiä erillisjulkaisuja 20 kpl (2 %). Nämä 910 julkaisua olivat siis sellaisia, joiden näkyvyyttä tarkasteltiin kansainvälisissä Scopus ja WoS –tietokannoissa.

Näkyvyys kansainvälisissä tietokannoissa oli seuraava (taulukot 13 ja 14): Scopuksesta löytyi 495 viitettä ja WoS:sta 373 viitettä. Pällekkäisiä viitteitä, jotka löytyivät kummastakin tietokannasta oli 59 % (321 kappaletta). Erilaisia VTT:n julkaisuja löytyi yhteensä 547 kappaletta. Tämä vastaa 60 % VTT:n 910:stä julkaisusta. Myös viittausten määrässä oli eroja. Scopuksesta löytyi selvästi enemmän viittauksia kuin WoS:sta.

VTT:n julkaisujen näkyvyys ja vaikuttavuus Scopuksessa ja WoS tietokannoissa		
	Scopus	WoS
Yleinen näkyvyys (löytyvyys tietokannoista)	495	373
Vaikuttavuus (saadut viittaukset tietokannoissa)	2003	1600

Taulukko 13. VTT:n 20 tutkijan julkaisujen näkyvyys ja vaikuttavuus Scopus- ja Web of Science -tietokannoissa.

VTT:n julkaisujen määrät ja osuudet: Samat, uniikit ja eri julkaisut Scopuksessa ja Web of Sciencessä				
Viitteitä	Scopus viitteet	WoS viitteet	Viitteet yhteensä ja osuus koko aineistosta	
Samoja viitteitä	321	321	321	59 %
Uniikkeja viitteitä	174	52	226	41 %
Yhteensä eri viitteitä	495	373	547	100 %

Taulukko 14. VTT:n 20 tutkijan julkaisujen samojen, uniikkien ja eri julkaisujen määrät ja osuudet Scopus- ja Web of Science -tietokannoissa.

Eri aiheryhmissä samojen viitteiden osuus Scopuksessa ja WoS:ssa vaihteli 42 ja 77:n %:n välillä (taulukko 15). Suurimmillaan se oli odotetusti biokemian alalla, jossa julkaistaan vakiintuneissa kansainvälisissä tieteellisissä lehdissä. Pienimmillään samojen viitteiden osuus oli tietokonetekniikan ja materiaalitekniikan aloilla. Julkaisukanavat näillä aloilla vaihtelevat enemmän ja mm. konferenssiesitelmiä oli aineistossa paljon.

VTT:n julkaisut aiheorymittäin: Samat julkaisut Scopusessa ja Web of Sciencessä					
Aiheryhmä (Scopus-luokitus)	Scopus- viitteet	WoS- viitteet	Yhteensä	Samoja viitteitä	Samojen viitteiden osuus %
Agricultural and biological sciences	27	22	30	19	63
Biochemistry, Genetics and Molecular biology	58	55	64	49	77
Computer Science	70	58	89	39	44
Energy	30	28	34	24	71
Engineering	73	52	73	52	71
Environmental Science	35	25	35	25	71
Materials Science	84	48	93	39	42
Physics and astronomy	118	85	129	74	57
Yht.	495	373	547	321	59 %

Taulukko 15. VTT:n 20 tutkijan julkaisut: samojen julkaisujen määrät ja osuudet eri aiheoryhmissä Scopus- ja Web of Science-tietokannoissa.

Uniikkimateriaalin määrässä ja osuudessa oli tietokannoissa selvä ero (taulukko 16). Scopusessa uniikkien viitteiden osuus oli 32 % ja vaihteluväli aiheorymittäin oli 14 – 48 %. WoS:ssa uniikkeja viitteitä on selvästi vähemmän, 10 %, ja vaihteluväli aiheorymittäin oli 0 – 21 %. Scopusessa oli eniten uniikkia materiaalia aiheoryhmissä Materials Science. WoS:ssa eniten uniikkia materiaalia oli aiheoryhmissä Computer Science, jossa myös Scopusen kohdalla oli runsaasti uniikkeja viitteitä. Tällä alalla julkaisukanavat ilmeisesti hajaantuvat laajalle.

VTT:n julkaisut aiheorymittäin: Uniikit julkaisut Scopusessa ja Web of Sciencessä					
Aiheryhmä (Scopus-luokitus)	Yhteensä viitteet	Uniikit, Scopus	Uniikit, Scopus %	Uniikit, WoS	Uniikit, WoS %
Agricultural and biological sciences	30	8	27	3	10
Biochemistry, Genetics and Molecular biology	64	9	14	6	9
Computer Science	89	31	35	19	21
Energy	34	6	18	4	12
Engineering	73	21	29	0	0
Environmental Science	35	10	29	0	0
Materials Science	93	45	48	9	10
Physics and astronomy	129	44	34	11	9
Yht.	547	174	32 %	52	10 %

Taulukko 16. VTT:n 20 tutkijan julkaisut: uniikkien julkaisujen määrät ja osuudet eri aiheoryhmissä Scopus- ja Web of Science-tietokannoissa.

Suurin ryhmä uniikkimateriaalia olivat konferenssit, joita Scopusessa on enemmän kuin WoS:ssa. Kuitenkin aineistossa oli myös paljon konferensseja, jotka olivat WoS:ssa, mutta eivät Scopusessa. Lisäksi kummankin tietokannan uniikkiaineistossa on runsaasti tieteellisten lehtien artikkeleita. Scopusen laajempi kansainvälinen kattavuus selittää osaksi uniikit artikkelit Scopusessa.

Esimerkkejä konferensseista/konferenssien järjestäjistä, joita tuli esille tässä VTT:n aineistossa ja joita Scopuksessa oli, mutta WoS:sta puuttui:

- Lecture Notes in Computer Science (ennen vuotta 2003 julkaistut)
- Lecture Notes in Artificial Intelligence (ennen vuotta 2003 julkaistut)
- Association for Computing Machinery (ACM): n konferensseja
- European Conference on Artificial Intelligence (ECAI) konferensseja
- SPIE:n, The International Society for Optical Engineering, konferensseja
- ASTM:n konferensseja
- IEEE:n konferenssisarjoja, esim.
 - o Computer Society of the Institute of Electrical and Electronics Engineering,
 - o International Conference on Robotics and Automation (ICRA)

Molemmissa tietokannoissa on selvää ja melko suurta horjuvuutta tieteellisten konferenssien ja artikkeleiden sisällyttämisessä. Sama sarja saattoi olla toisena vuotena toisessa ja seuraavana vuonna toisessa, mutta puuttua toisesta. Ilmeisesti tietokannoissa on melko paljon virheitä, koska jo näin pienessä otoksessa horjuvuus tulee ilmi. Selvästi esille tuli myös erilainen tapa kirjata konferensseja, joko konferenssin nimen tai järjestäjän mukaan.

Viittauksia on lähes kautta linjan hieman enemmän Scopuksessa kuin WoS:ssa, mutta muutamassa tapauksessa myös päinvastoin. Keskeinen syy Scopuksen suurempaan viittausmäärään lienee se, että viittaukset tulevat Scopuksen suuremmasta julkaisumassasta. Esimerkiksi yllämainitut tunnetut konferenssisarjat, ovat sellaisia, joihin viitataan paljon.

Taulukossa 17 on esitetty aiheryhmittäin tutkijakohtaisesti julkaisutuotannon saamat viittaukset ja niiden perusteella lasketut Hirsch-indeksit Scopuksessa ja WoS:ssa. Hirsch-indeksi näytti tasoittavan tietokantojen eroja viitteiden ja viittauksien määrässä. Hirsch-indeksit olivat hyvin lähellä toisiaan, mutta kuitenkin ne ovat Scopuksessa hieman korkeampia kuin WoS:ssa.

Julkaisut/tutkija aiheryhmittäin ja julkaisut/niiden saamat viittaukset sekä Hirsch-indeksit Scopus- ja WoS-tietokannoissa			
VTT:n tutkijan julkaisujen Scopus-aihe ryhmä		Scopus-Hirsch-indeksi	WoS- Hirsch-indeksi
Agricultural and biological sciences	Tutkija 1	8	5
Biochemistry, Genetics and Molecular biology	Tutkija 1	12	11
	Tutkija 2	5	5
Computer science	Tutkija 1	3	1
	Tutkija 2	6	4
	Tutkija 3	3	3
	Tutkija 4	5	4
Energy	Tutkija 1	3	2
	Tutkija 2	2	2
Engineering	Tutkija 1	5	6
	Tutkija 2	1	1
	Tutkija 3	1	1
	Tutkija 4	4	3
Environmental Science	Tutkija 1	6	6
Materials science	Tutkija 1	2	
	Tutkija 2	4	5
	Tutkija 3	3	2
	Tutkija 4	4	4
Physics and astronomy	Tutkija 1	8	7
	Tutkija 2	8	7

Taulukko 17. VTT:n 20 tutkijan julkaisut aiheryhmittäin ja niiden saamat Hirsch-indeksit Scopus- ja Web of Science-tietokannoissa.

4.3. Tietotekniikan alan konferenssit

Konferenssit ovat tärkeä teknistieteellisen tutkimuksen julkaisufoorumi. Esimerkiksi tietojenkäsittelytieteessä tieteelliset konferenssijulkaisut ovat ns. referoituihin aikakauslehtiin verrattavissa oleva alkuperäistulosten julkaisukanava.

Julkaisutietokannoissa konferenssiartikkelien ja –esitysten näkyvyys on usein huonompi kuin lehtiartikkelien. Tämä antaa asianomaisten tutkimusalojen laatutasosta harhaanjohtavan heikon kuvan. Professori Esko Ukkosen ja tietotekniikan tutkimuslaitoksen HIIT työryhmä tarkastelivat kongressijulkaisemisen asemaa ja laatukriteerejä osana Opetusministeriön KOTA-tietokannan kehittämiseen liittyvää työtä syksyllä 2006.

Esko Ukkosen johdolla työryhmä kokosi esimerkinomaisen luettelon tietotekniikan alan konferensseista (liite 1) ja totesi, että laatukongressien vahvistettua luetteloa on tuskin mahdollista ylläpitää, sillä nopeasti etenevillä aloilla tilanne muuttuu koko ajan. Lisäksi konferenssien luokittelu tulee käytännössä tehdä tapauskohtaisesti yleisten kriteerien pohjalta.

Laatukongressin peruskriteereiksi ehdotettiin seuraavia tekijöitä:

- konferenssin tulee olla kansainvälinen, vakiintunut ja säännöllisesti järjestettävä
- konferenssin artikkelit valitaan vertaisarviointimenettelyllä, joka karsii huomattavan osan tarjotuista papereista, artikkeleiden sallittu maksimipituus on vähintään 6 sivua ja artikkelit on julkaistu konferenssijulkaisussa, joka on yleisesti saatavissa konferenssin jälkeen
- konferenssista vastaa tunnettu tieteellinen organisaatio tai konferenssijulkaisulla on tunnettu kansainvälinen kustantaja

Tässä esiselvityksessä tarkasteltiin Ukkosen työryhmän esimerkkeinä ilmoittaman tietotekniikan alan kansainvälisten ja vakiintuneiden konferenssien näkyvyyttä Scopus ja ISI Proceedings tietokannoissa. Konferenssiartikkeleiden saamat viittaukset indeksoidaan Scopus-tietokantaan, mutta ei ISI Proceedings tietokantaan.

Kaikkiaan Ukkosen työryhmän luettelossa olleista tietotekniikan alan liitteessä 1 esitetyistä 40 konferenssista

- 26 (65 %) oli indeksoitu Scopus tietokantaan ja
- 22 (55%) ISI Proceedings tietokantaan.

Tulos osoittaa, ettei konferenssijulkaisujen löytyvyys ole hyvä kummassakaan tietokannassa.

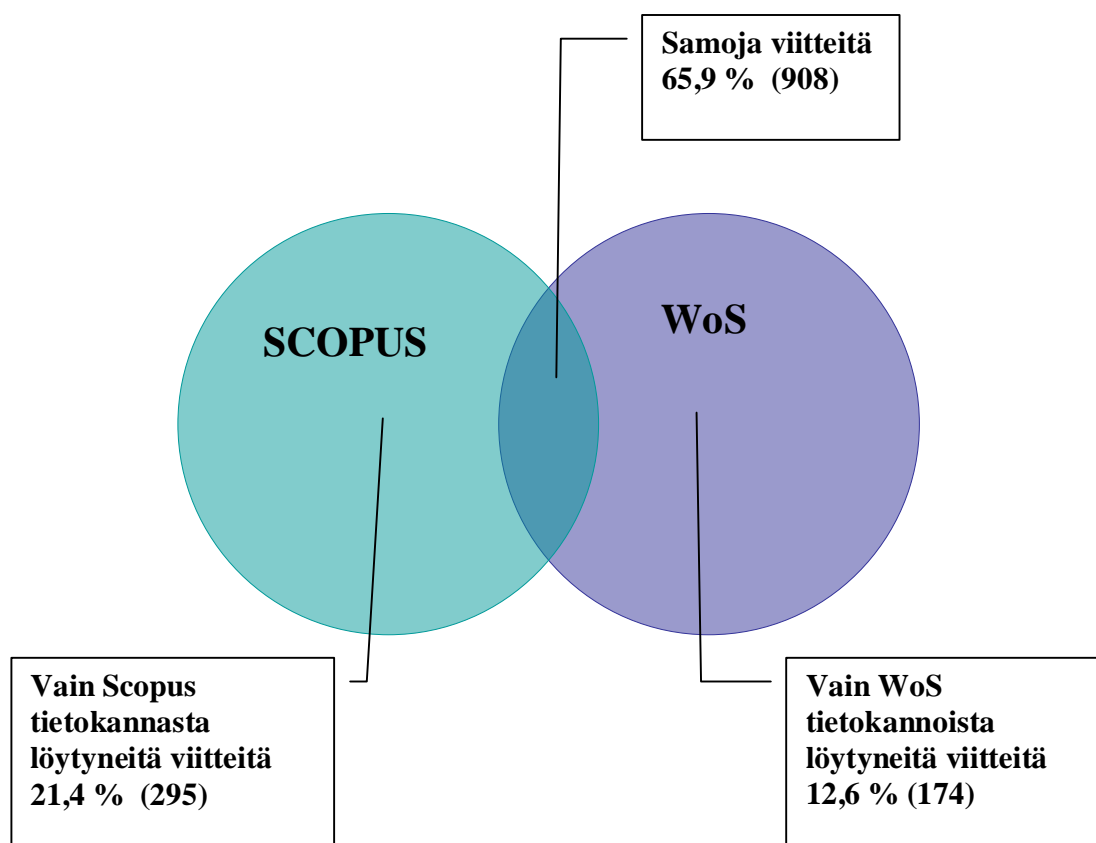
Tiedeyhteisön oman järjestön konferenssi voi olla säännöllisesti järjestettävä ja vakiintunut, mutta konferenssin julkaisusta vastaa usein vuosittain vaihtuva paikallinen järjestäjä. Tällöin konferenssijulkaisun nimi ja muut tunnistetiedot vaihtuvat joka vuosi eikä säännöllisesti ilmestyvällä julkaisulla ole ISSN-sarjanumeroa. Tämä heikentää konferenssijulkaisujen saatavuutta ja näkyvyyttä tietokannoissa oleellisesti. Tämä heijastui Ukkosen työryhmän aineistossa siten, että esimerkiksi sellaisten suurten tiedeorganisaatioiden kuin ACM ja IEEE tukemien konferenssien julkaisujen tiedoista osa puuttui sekä Scopus- että ISI Proceedings tietokannoista. Nämä molemmat tiedeorganisaatiot julkaisevat konferenssijulkaisunsa myös verkkopalvelussaan, joten tieto em. julkaisuista on kuitenkin olemassa. Hämmennystä aiheutti myös se, että Springer-kustantajan Lecture Notes in Computer Science (LNCS) –sarjassa ilmestyneet konferenssijulkaisut löytyivät osittain myös Science Citation Index –tietokannasta, jolloin LNCS julkaisujen saamat viittaukset olivat tältä osin saatavissa.

Myös konferenssien jatkuvassa seurannassa oli puutteita: ISI Proceedings tietokannassa 14:stä jatkuvasti ilmestyvästä konferenssijulkaisuista löytyi tiedot vain muutaman vuoden osalta.

5. Johtopäätökset ja julkaisuihin perustuvien arviointien ajankohtaisuudesta

5.1. Esiselvityksen tulosten osviitat

Yhteensä tarkastelussa oli mukana 1377 Scopus tai WoS tietokannoista löytynyttä TKK:n tai VTT:n tutkijoiden kansainvälistä teknistieteellistä julkaisua. Esiselvityksen tulokset (kuva 3) osoittavat, että kaksi kolmasosaa, 66 %, Scopusesta ja WoS:ta löytyneistä viitteistä on samoja. Vain Scopus tietokannoissa olleita viitteitä oli kuitenkin merkittävästi enemmän, yhteensä 295 kpl, mikä on 21 % kaikista viitteistä. Vain WoS tietokannoissa olleita viitteitä oli 13 % (174) kaikista viitteistä.



Kuva 3. Tarkastelussa olleen TKK:n ja VTT:n aineiston, yhteensä 1377 julkaisua näkyvyys Scopus ja WoS tietokannoissa.

Noin 34 % löytyneestä aineistosta on uniikkia eli samaa viitettä ei löydy toisesta tietokannasta. Uniikkien viitteiden määrä on selvästi korkeampi Scopuksessa kuin WoS:ssa. Se on Scopuksessa koko aineistossa 21,4 % ja WoS:ssa 12,6 %. Tässä oli kuitenkin merkittävä ero TKK:n ja VTT:n aineistoissa: TKK:n aineistossa uniikkia aineistoa löytyi saman verran sekä Scopuksesta että WoS tietokannoista (taulukko 9) kun taas VTT:n aineistossa uniikkien viitteiden määrä Scopuksessa oli huomattavasti korkeampi, 32 % kaikista ja vastaavasti WoS osuus jäi pienemmäksi, 10 % (taulukko 16).

- *Toisin sanoen esiselvityksen tulosten valossa: mikäli tehdään tiedonhaku pelkästään WoS:sta, jää hieman yli 20 % viitteistä saamatta Scopukseen verrattuna. Jos tiedonhaku tehdään Scopuksesta, jää saamatta vain vajaa 13 % viitteistä WoS:iin verrattuna.*

Samanlaiseen lopputulokseen päätyi myös Ulf Kronman selvittäessään Kungliga Tekniska Högskolanin, KTH:n, julkaisutuotannon näkyvyyttä Scopuksessa ja WoS:ssa (Kronman 2007). KTH:n julkaisujen näkyvyys oli noin 20 % parempi Scopuksessa kuin Web of Sciencessä.

Myös erot viitausten määrässä ja päällekkäisyydessä vastaavat aikaisempia julkaistuja selvityksiä. Bosman et al.:n selvityksessä [Bosman et al. 2006] tarkasteltiin viittaavien artikkeleiden päällekkäisyyttä. Scopuksen ja WoS:n eroista päädyttiin pienehkön otoksen perusteella siihen, että useimmissa aihepiireissä Scopuksen ja WoS:n erot ovat suhteellisen pieniä, viittaavista artikkeleista 80 – 90 % ovat samoja.

Horjuvuus dokumenttien sisällyttämisessä tietokantoihin, mikä tuli tässä TKK:n ja VTT:n esiselvityksessä esille, on havaittu myös aikaisemmissa selvityksissä. Esimerkiksi Kronman [2007] mainitsee siitä KTH:n julkaisujen näkyvyyttä koskevassa esityksessään. Hänen mukaansa kysymys on laajojen tietokantojen ongelmasta luoda mekanismeja, joilla tavoittaa kaikki halutut julkaisut. Ongelmana ovat esimerkiksi nimiään vaihtelevat konferenssarjat.

Mehon and Yangin [2007] Scopuksen ja WoS:in eroja käsitelleen selvityksen yksi lopputulemista oli, että kun halutaan mahdollisimman hyvä kuva tutkijoiden kansainvälisestä kommunikaatiosta ja vaikuttamisesta on hyvä käyttää toisiaan täydentäviä lähteitä. Kun pitäydytään pelkästään tieteellisten lehtien tarkastelussa, tieteellisen kanssakäymisen kuva jää vajaaksi, koska joillakin aloilla vaikuttamisen pääasialliset foorumit ovat muita. Scopuksen laajempi eri dokumenttityyppien kirjo ja maantieteellinen kattavuus antaa olennaista tietoa tutkijoiden välisestä tieteellisestä kommunikaatiosta ja siksi se voi tarjota hyvän lähtökohdan tai lisäaineiston vaikuttavuuden selvittämiseen kattamallaan tieteenaloilla.

Esiselvityksen perusteella voisi todeta, että kun tehdään bibliometrisia analyysyjä, täytyy voida perustella, miksi valitaan tietty tietokanta työvälineeksi. Tietokantojen paremmuus riippuu kysymyksessä olevasta tieteenalasta ja sen tieteellisen kommunikaation käytänteistä. Valittu tietokanta merkitsee aina valittua näkökulmaa selvitetävään asiaan.

Kun tehdään esimerkiksi organisaatioiden, tieteenalojen tai maiden välisiä vertailuja, tulisi näkökulman olla mahdollisimman laaja.

5.2. "Measuring Science"- kurssi

Leidenin yliopiston tieteen ja teknologian tutkimuskeskus CWTS järjestää vuosittain tutkimustoiminnan arviointiin liittyvän viikon intensiivikurssin "Measuring Science". Irma Pasanen TKK:n kirjastosta ja Pirjo Sutela VTT Tietoratkaisuista osallistuivat kurssille 8.-12.10.2007. Kurssille osallistui noin 30 tutkijaa, hallintovirkamiestä tai tietoasiantuntijaa pääosin eurooppalaisista yliopistoista ja tutkimuslaitoksista. Kaukaisimmat osallistujat olivat Etelä-Koreasta.

CWTS toimii Hollannissa Leidenin yliopiston yhteiskuntatieteiden tiedekunnan alaisena tutkimuslaitoksena ja se on erikoistunut vaativiin tieteen suoritusta arvioiviin ja organisaatorakenteita koskeviin tutkimuksiin. Laitos tekee arviointeja niin EU:n, valtioiden hallitusten, tutkimuslaitosten ja yliopistojen kuin yritystenkin toimeksiannosta. Bibliometrinen mittausytimenä toimivat Thomson Scientific ISI tietokannat, jotka CWTS on hankkinut omiin tietojärjestelmiinsä. Tutkimuskeskus on myös käynnistämässä laajahkoa selvitystä Scopus-tietokannan soveltuvuudesta bibliometrisiin mittauksiin. Kurssilla perehdyttiin lisäksi lyhyesti bibliometrisiin karttoihin, joiden avulla voidaan visualisoida esimerkiksi tutkimustoiminnan aktiivisuutta, suuntautumista, yhteistyötä tms. (<http://www.cwts.nl/>)

bibliometrisissa arvioinneissa tunnetaan käsite "Crown indicator" eli suhteellisen vaikuttavuuden ilmaisin. Eri alojen julkaisu- ja viittauskäytänteiden erot otetaan huomioon tätä ilmaisinta laskettaessa. CWTS:n tutkimuksissa on kuitenkin voitu lisäksi osoittaa, että yksikön koolla on tästä huolimatta merkitystä myös näin saadun vaikuttavuuden arvoon. Pienen yksikön tai tieteen alan on helpompi saavuttaa korkeampi taso kuin suuren yksikön. CWTS on kehittänyt oman, suhteellisen vaikuttavuuden ilmaisimeen perustuvan ranking-menetelmänsä, ja julkaisee tähän perustuvaa eurooppalaisten huippuyliopistojen listaa.

5.3. Ranking

Kiristyvän kilpailun myötä erilaiset yliopistojen ja tutkimusyksiköiden kansainvälistä paremmuusjärjestystä kuvaavat luettelot ovat yleistyneet 2000-luvulla. Tutkimusjulkaisuja ja niiden saamia viittauksia käytetään näissä arvioinneissa yhtenä mittarina.

ARWU, Academic Ranking of World Universities, Shanghai Jiao Tong yliopiston tuottama lista, käyttää lähteenään ISI Thomson Scientificin tietokantoja seuraavin painoarvoin:

- WoS tietokannoissa olevat julkaisut 20 %
- ISI Highly Cited tutkijoiden lukumäärä 20 %

Lisäksi julkaisuja koskevana indikaattorina käytetään Science- ja Nature lehdissä ilmestyneitä julkaisuja (painoarvo 20%). Muut indikaattorit ovat Nobel-palkinnot (yliopiston henkilökunta painoarvolla 20 % ja alumni 10%) sekä opettajien ja tutkijoiden lukumäärä (FTE, Full time equivalent) 10 %. (ARWU 2007)

THES, The Times Higher Education Supplement siirtyi vuonna 2007 käyttämään SCOPUS tietokantaa julkaisujen saamien viittausten lähteenä. Saadut viittaukset per tutkijoiden lukumäärä saa kokonaisarvioinnissa painoarvon 20 %. Muut indikaattorit ovat kansainvälisten asiantuntijoiden vertaisarviointi (painoarvo 40 %), opiskelija per opettaja suhde (painoarvo 20 %), kansainvälisten opiskelijoiden (painoarvo 5 %), ja kansainvälisten opettajien ja tutkijoiden (painoarvo 5 %) osuus sekä valmistuneita opiskelijoita rekrytoivien tahojen antamat arviot (10 %).

THES käytti aikaisemmin julkaisuja koskevassa aineistossa ISI Thomson Scientificin ESI-tietokantaa. Johtuen alakohtaisista kynnsarvoista jää moni tutkimuslaitos kokonaan huomioonottamatta ESI-tietokannassa. Siirtymistä SCOPUS-tietokantaan julkaisuja ja niiden saamia viittauksia koskevan tiedon hankinnassa on perusteltu lisäksi mm. seuraavilla syillä (THES 2007):

1. Scopus painottuu vähemmän Yhdysvalloista lähtöisin olevaan aineistoon
2. Scopus kattaa laajemman määrän artikkeleita ja aikakauslehtiä, jolloin vähemmän tunnettujen yliopistojen ja vähemmän julkaisevien korkeakoulujen näkyvyys paranee
3. Scopus kattaa enemmän muulla kuin englannin kielellä julkaistuja aineistoja

Green Ranking on Leidenin yliopiston tieteen ja teknologian tutkimuskeskus CWTS:n tuottama 100 suurinta eurooppalaista yliopistoa koskeva paremmuusjärjestystä kuvaava luettelo, joka perustuu 100 % tutkimusjulkaisuihin (Enserink 2007). Tietopohjana käytetään ISI Thomson Scientific tietokantoja, jotka CWTS on hankkinut omiin tietojärjestelmiinsä. Ranking perustuu julkaisujen saamiin viittauksiin (keskimääräiseen vaikuttavuuteen), ja laskelmissa on huomioitu yliopiston koko ja tieteenalakohtaiset erot.

6. Esiselvityksen tuloksen arviointi

Esiselvitys, jossa analysoitiin TKK:n ja VTT:n tutkijoiden näkyvyyttä ja vaikuttavuutta Scopus ja Web of Science –tietokannoissa ja tutustuttiin aihepiiriä käsittelevään kirjallisuuteen, osoitti että:

- suuri osa Scopusesta ja WoS:ta löytyneistä viitteistä on samoja: esiselvityksen mukaan 66 %
- molemmissa tietokannoissa on uniikkia aineistoa, Scopusessa enemmän:
 - vain Scopus tietokannassa oli 21,4 % kaikista viitteistä
 - vain WoS tietokannoissa oli 12,6 % kaikista viitteistä.
- vaikuttavuus (viittausten määrä samoilla julkaisuilla) oli suurempi Scopus – tietokannassa
- Hirsch-indeksi oli hieman parempi Scopus –tietokannassa
- konferenssijulkaisujen näkyvyys ei ollut kattava kummassakaan

Molemmat tietokannat osoittautuivat arvokkaiksi lähteiksi näkyvyyden ja vaikuttavuuden arvioinnissa. Tulokset kuitenkin osoittavat, että Scopus ja Web of Science –tietokannoista saaduissa *tuloksissa on eroa Scopusen eduksi ainakin teknisillä aloilla*. Siitä syystä olisi tärkeää huomioida Scopusen mahdollisuudet ja lisäarvo laajoja näkyvyyden ja vaikuttavuuden arvioita tehtäessä. Yksi mahdollisuus selvittää tarkemmin Scopusen käyttömahdollisuuksia, olisi kokeilla Suomessa uutta bibliometrisia mittauksia varten kehitettyä tuotetta, Scopus Custom_dataa. Se tarjoaa käyttäjälle räätälöityjä otteita Scopus -tietokannasta XML-muodossa ja siihen kuuluu myös analysointiohjelmisto.

Lisäksi esiselvitys osoitti, että osa julkaisuista ei saanut lainkaan näkyvyyttä Scopus ja WoS tietokannoissa. Noin 40 % TKK:n ja VTT:n esiselvityksen tarkasteluun valitusta julkaisuista jäi kokonaan löytymättä Scopus ja WoS tietokannoista. Nämä olivat suureksi osaksi kansainvälisiä konferenssijulkaisuja.

Esiselvityksessä ei tarkasteltu laajemmin tieteenalakohtaisia eroja julkaisutoiminnassa tai viittauskäytännöissä. Nämä seikat vaikuttavat kuitenkin merkittävästi saavutettavaan näkyvyyteen ja vaikuttavuuteen. Kun ISI Thomson Scientificin tietokantojen tai Scopusen pohjalta arvioidaan tutkimuksen tuloksellisuutta, on tuloksia syytä tarkastella ja esittää huolella. Lopputulos on tulkittava käytetyn lähdeaineiston valossa ja sen sisältämän luokituksen mukaisesti.

Lähteet

ARWU: Ranking Methodology. 2007.

<http://www.arwu.org/rank/2007/ARWU2007Methodology.htm>

Ball, R., Tunger, D. 2006. Science indicators revisited - Science citation index versus SCOPUS: A bibliometric comparison of both citation databases .Information Services and Use, 26 (4), pp. 293-301.

Bosman, J., van Mourik, I. , Rasch, M. , Sieverts, E. and Verhoeff, H. 2006. Scopus reviewed and compared: The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar. eprints.rclis.org/11966/.

Dess, H. M. 2006. Database Reviews and Reports. Scopus. Issues in Science and Technology Librarianship, Winter 2006.

Enserink, M. 2007. Who Ranks the University Rankers? Science, 317 (5841), pp. 1026-1028. www.sciencemag.org

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 2005 November 15; 102(46), pp. 16569–16572. Published online 2005 November 7. doi: 10.1073/pnas.0507655102.

Kronman, U. 2007. Bibliometrics – what, why and how? 5th Elsevier Scandinavian Librarian Forum. 8th – 9th November 2007. Oslo, Norway.

Meho, L.I., Yang, K. 2007. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. Journal of the American Society for Information Science and Technology, 58 (13), pp. 2105-2125.

Moed, H. F. 2005. Citation Analysis in Research Evaluation. [Dordrecht (Netherlands)]: Springer. 346 p. ISBN-13 978-1-4020-3713-9.

Moya-Anegón, F. De, Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F.J., González-Molina, A., Herrero-Solana, V. 2007. Coverage analysis of Scopus: A journal metric approach. Scientometrics, 73 (1), pp. 53-78.

THES: Basic explanation of key enhancements in methodology for 2007. <http://www.topuniversities.com/>

LIITE 1

AAAI	National Conference on Artificial Intelligence
CCC	IEEE Conference on Computational Complexity
COLT	Computational Learning Theory (COLT)
CRYPTO	Advances in Cryptology
DOA	International Symposium on Distributed Objects and Applications
Ecoop	European Conference on Object-Oriented Programming (ECOOP)
EDOC	Enterprise Distributed Object Computing
FOCS	IEEE Symposium on Foundations of Computer Science
ICALP	International Colloquium on Automata, Languages and Programming (ICALP)
ICDM	IEEE International Conference on Data Mining (ICDM)
ICFP	ACM SIGPLAN International Conference on Functional Programming
ICML	International Conference on Machine Learning
ICWE	International Conference on Web Engineering
IEEE ICSE	ACM IEEE ICSE International Conference on Software Engineering
IFIP/ACM	
Middleware	ACM/IFIP/USENIX International Middleware Conference
IJCAI	International Joint Conference on Artificial Intelligence
INFOCOM	IEEE Conference on Computer Communications
ISMB	Annual International Conference on Intelligent Systems for Molecular Biology
ITC	
Specialist Seminar	International Teletraffic Congress
IWQoS	IEEE International Workshop on Quality of Service
KDD	ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
Mobicom	ACM International Conference on Mobile Computing and Networking
Mobihoc	ACM International Symposium on Mobile Ad Hoc Networking and Computing
NIPS	Neural Information Processing Systems
PODC	ACM Symposium on Principles of Distributed Computing
PODS	ACM Symposium on Principles of Database Systems
POPL	ACM SIGPLAN Principles of Programming Languages
RECOMB	Research in Computational Molecular Biology
SIGGRAPH	ACM SIGGRAPH: International Conference on Computer Graphics and Interactive Techniques
SIGIR	ACM Conference on Research and Development in Information Retrieval
SIGMOD	ACM International Conference on Management of Data
SIGPLAN	ACM SIGPLAN Object Oriented Programming Systems Languages and Applications Conference
OOPSLA	
SIGSOFT	
FSE	ACM SIGSOFT Foundations of Software Engineering
SoCG	SCG ACM Annual Symposium on Computational Geometry
SODA	ACM-SIAM Symposium on Discrete algorithm
STACS	Symposium on Theoretical Aspects of Computer Science (STACS)
STOC	ACM Symposium on Theory of Computing
UAI	Conference on Uncertainty in Artificial Intelligence
VLDB	Very Large Data Bases (VLDB) Conference
WWW	ACM International World Wide Web Conference