

METSÄTIETEELLISEN KIRJALLISUUDEN HAKU ATK-TIEDOSTOISTA

ANNIKKI KARJALAINEN, SINI KÄRKI-FOLKERSMA & LAURI HYVÄKKÄ

Summary

EXPERIENCES OF SOME BIBLIOGRAPHIC DATABASES IN FORESTRY

Saapunut toimitukselle 17. 6. 1983

Selvityksessä tarkastellaan eräiden kansainvälisten viitetietokantojen soveltuvuutta suomalaisten metsäntutkijoiden tarpeisiin. Aineistona oli 26 metsäntutkijalle tehdyt ilmaiset haut, joiden tulokset analysoitiin arviointilomaketta käyttäen.

Tietokonepohjainen kirjallisuushaku oli 62 %:lle tutkijoista ensimmäinen. Hauissa käytettiin pääasiassa AGRICOLA-, BIOSIS- ja CAB -tietokantoja. Käytetyin tietokanta oli CAB, johon sisältyy mm. metsätieteellinen Forestry Abstracts-osatietokanta. CAB:n tarkkuus oli 51,0 %. AGRICOLA -tietokannasta saatiin edellistä vähemmän viitteitä, mutta tarkkuus oli parempi, 58,3 %. BIOSISin tarkkuus oli edellisiä heikompi, 35,2 %, kuitenkin esim. ekologisissa aiheissa hyvä. Kustannukset olivat keskimäärin 260 mk hakua kohti. AGRICOLAn käyttö oli edullisinta.

Tulosten päällekkäisyys oli varsin pieni, joten hakujen suorittaminen useammasta tietokannasta on perusteltua kustannusten noususta huolimatta. Vajaa puolet tutkijoista piti haun tulosta hyvänä tai melko hyvänä ja lähes kaikki olivat halukkaita käyttämään tulevaisuudessa maksullisia palveluja. Käsitkset kustannusten kohtuullisuudesta vaihtelivat.

Selvityksessä esitetään myös katsaus edellä mainituista tietokannoista aiemmin tehtyihin vertaileviin tutkimuksiin. Lopuksi tarkastellaan metsätieteellisen tiedonvälityksen puutteita ja kirjallisuushaun onnistumiselle tärkeitä tekijöitä.

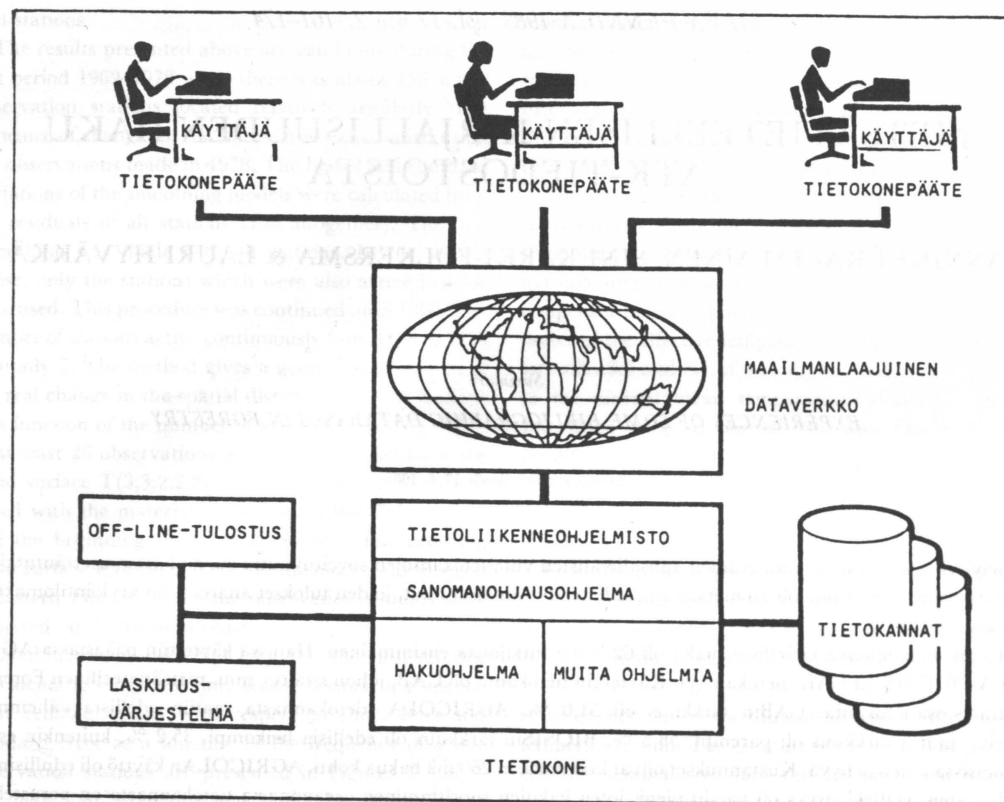
1. TIETOKONEPOHJAISET TIEDONHAKUJÄRJESTELMÄT

Luonnontieteissä, lääketieteessä ja tekniikassa on tietojen tallentamiseen ja välittämiseen käytetty tietokoneiden muistia jo 1960-luvun alusta lähtien. Tekniikan kehittyessä on syntynyt tietopalvelujärjestelmiä, joiden osina ovat tietokoneet tietokantoinen ja tietoliikenneverkostot. Kaapeli- ja satelliittiyhteyksin ne tekevät tietokantojen käytön mahdolliseksi pitkienkin etäisyyksien päästä. Kuvassa 1 on esitetty tällaisen online- eli kyselykäyttöisen tiedonhakujärjestelmän osat.

Tietopankkien käytön yleistymiseen on vaikuttanut ratkaisevasti ns. pakettikytkentäisten tiedonsiirtoverkkojen kehittyminen. Yhdysvalloissa toimivilla TYMNET ja TE-

LENET -verkoilla, joihin sikäläiset tiedonhakujärjestelmät on kytketty, on solmukohtia myös Euroopassa. Euroopassa puolestaan toimii EURONET ja Skandinaviassa SCANET (Laitinen 1981).

Tietopankkeihin tallennettu tieto on erillisinä tietokantoina. Faktatietokannoista itse tieto voidaan saada suoraan, esim. tilastot, ennusteet ym. Suurin osa on kuitenkin viitetietokantoja eli bibliografisia tietokantoja, jotka toimivat hakemistoina eri tavoin julkaistuun tietoon. Haun tuloksena saadaan tietokonepäätteelle viittaus aikauslehtiartikkeliin, tutkimusraporttiin tai virallisjulkaisuun.



Kuva 1. Online-informaatiopalvelujärjestelmän osat (SDC Search Service, esite).

Tiedonhaku viitetietopankista tapahtuu valitsemalla ensin sopiva tietokanta ja antamalla hakusana. Järjestelmä ilmoittaa, montako viitettä ko. hakusanalla löytyy. Tämän jälkeen käyttäjä antaa lisää hakusanoja ja lopuksi hakusanojen välisen loogisuusyhtälön, jolloin tietokannasta voidaan tulostaa ne viitteet, jotka todennäköisimmin antavat vastauksen esitettyyn kysymykseen (vrt. Henry 1980, Lancaster 1979).

Kirjallisuusviitetietopankissa kustakin dokumentista on esitetty ns. bibliografiset tiedot, joiden perusteella dokumentti voidaan identifioida. Tietoa voidaan hakea näiden eri tietoalkioiden eli tekijän nimen, lehden nimen ja artikkelin otsikkosanojen ym. perusteella. Lisäksi on valittu muutamia artikkelin sisältöä kuvaavia indeksitermejä eli hakusanoja, jolloin tiedonhaku voidaan perustaa niihin samoin kuin painettujen julkaisujen asiahakemistoissa.

Jotta viitteiden perusteella voitaisiin paremmin päätellä, kannattaako alkuperäiseen dokumenttiin tutustua, on kirjallisuusviitetietokantoihin usein myös syötetty lyhyt tiivistelmä dokumentista. Tiivistelmän sanoja voidaan käyttää hyväksi niinikään tiedonhaussa.

Tiedonhaun kustannukset koostuvat tietojärjestelmän käytöstä aiheutuvista kustannuksista ja tietoliikennekustannuksista. Tietokantojen käyttöveloitukset vaihtelevat 130–yli 600 markkaan tunnilta. Tietoliikennekustannukset vaihtelevat tietokoneen sijaintipaikasta ja tiedonsiirtoverkosta riippuen. Tavallinen tiedonhaku, joka kestää 10–20 min. maksaa 50–600 mk.

Tiedonhaun tuloksena saadaan usein joukko viitteitä tietyistä julkaisuista. Julkaisun tyyppistä ja sijaintipaikasta riippuen se voidaan saada käyttöön ostamalla, lainaamalla tai hankkimalla siitä jäljenne.

Helsingin yliopiston metsäkirjastossa on

tehty tietokonepohjaisia kirjallisuushakuja vuodesta 1978 lähtien. Lockheed Dialog (myöhemmin Dialog Information Services) ja System Development Corporation -automatoiduista informaatiopalvelujärjestelmistä voidaan kirjallisuusviitteitä ja asiatietoja hakea yli 300 eri tieteenaloja edustavasta tietokannasta. Näistä Forestry Abstracts ja Forest Products Abstracts ovat käytetyimpiä metsätieteiden tietokantoja (sisältyvät laajaan CAB-tietokantaan). Metsätieteiden soveltavan luonteen vuoksi metsätutkijoita kiinnostavaa informaatiota saadaan myös muiden tieteenalojen tietokannoista (taulukko 1).

Suomen metsätutkijoiden tiedontarpeista ja tiedonsaantikanavista on tehty kaksi suppeahkoa haastattelua ja kyselytutkimusta (Hyväkkä 1978 ja Suckcharoen 1979). Atk-palvelujen kehittämiseksi Metsäkirjastossa pidettiin tärkeänä kokeilla käytettävissä olevia tietokantoja ja arvioida hausta saadut tulokset. Samalla haluttiin selvittää tiedon tarpeisiin ja saantiin liittyviä ongelmia. Tätä tehtävää varten saatiin Suomen Akatemialta Suomen Metsätieteellisen Seuran kautta syksyllä 1981 apuraha. Tutkimusprojektin vastuullisena johtajana toimi kirjastonhoitaja Annikki Karjalainen ja työryhmän muina jäseninä MH Lauri Hyväkkä ja tutkimusavustaja LuK Sini Kärki-Folkersma, joka työsti tutkimustulokset raportin muotoon.

Käsikirjoituksen ovat lukeneet MMT, FT Erkki Annila ja prof. Matti Kärkkäinen. Heille, samoin kuin kaikille muille, jotka ovat myötävaikuttaneet selvityksen valmistamiseen, esitämme parhaat kiitoksemme.

2. AIKAISEMMAT TUTKIMUKSET

2.1. Viitteiden saanti ja tarkkuus

Saaduista kirjallisuusviitteistä ja dokumenteista ovat relevantteja ne, jotka vastaavat annettua aihetta. Saannin eli löytyvyyden määräämiseksi verrataan atk-hausta saatujen relevanttien viitteiden määrää kaikkiin kyseisessä tietojärjestelmässä olevien relevanttien viitteiden määrään. Usein käytetty kriteeri informaatiopalvelujen onnistumisen mittauksessa on myös tarkkuus. Tarkkuus tarkoittaa relevanttien dokumenttien määrää verrattuna kaikkien saatujen dokumenttien määrään (Henry 1980, Lancaster 1979). Seuraavissa vertailuissa on useimmiten jouduttu tyytymään eri tietokannoista saatujen relevanttien viitteiden määrien vertailuihin.

Brugge ym. (1979) ovat tehneet vertailevan tutkimuksen eri tietokannoista PUDOCille (Centre for Agricultural Publishing and Documentation). Tutkimuksessa haettiin tietoa eri tietokannoista seitsemään tyypilliseen metsätieteelliseen kysymykseen ja viiteen kysymykseen, joiden aiheet olivat metsätieteiden reuna-alueilta. Haut tehtiin 21 tietokannasta, jotka jaettiin kolmeen luokkaan:

1. Metsätieteellinen informaatio on merkittävä tietokannan osa: CAB (sisältää Forestry Abstractsin), AGRICOLA, AGRIS, USDA/CRIS
2. tallentavat informaatiota muista aiheista: BIOSIS, COMPENDEX, CHEMICAL ABST-

RACTS, ENVIROLINE, EPB, EXCERPTA MEDICA, INSPEC, LABODOC, SOCIAL SCI-SEARCH

3. tallentavat tieteidenvälistä informaatiota: COMP, CONFERENCE PAPERS, NTIS, PASCAL, SSIE, SCISEARCH

Forestry Abstracts on yksi CABin noin 30 osatietokannasta. Forestry Abstracts -tiedosto antoi eniten relevantteja viitteitä. Tarkkuus oli kuitenkin vain 37 %. Seuraavina olivat AGRICOLA, BIOSIS ja PASCAL.

Metsätieteiden reuna-alueilta oleviin kysymyksiin saatiin viitteitä hyvin laajalti eri tietokannoista. Em. viiden kysymyksen kannalta tärkeimmät tietokannat olivat CAB, AGRICOLA, ENVIROLINE, SOCIAL SCI-SEARCH, NTIS ja PASCAL. Kysymyksestä riippuen ne antoivat joskus enemmän informaatiota kuin Forestry Abstracts.

Seitsemään tyypilliseen metsätieteiden kysymykseen saaduista relevanteista viitteistä oli 45 % Forestry Abstracts-osatietokannasta ja 55 % koko CABista. AGRICOLAssa ja AGRISissa oli noin 20 % relevanteista viitteistä, joita ei ollut muissa tietokannoissa. Erikoistuneet tietokannat, kuten BIOSIS ja ENVIROLINE antoivat 10–20 % kaikista relevanteista viitteistä ja monitieteelliset tietokannat, kuten PASCAL, NTIS ja SCISEARCH noin 15–25 %.

Viiteen ei-tyypilliseen metsätieteiden kysymykseen Forestry Abstracts, AGRICOLA ja

Taulukko 1 Eräitä metsätieteitä sisältäviä tietokantoja, 1981.

Tietokanta	Painettu muoto	Tuottaja	vuodesta	viitteitä	Aihealue	Kattavuus
AGRICOLA	Bibliography of Agriculture	National Agricultural Library (NAL), USA	1970-	1 500 000	maatalous	kirjoja, sarjajulk., kongressijulk., raportteja, FAO:n ja USDA:n julk. ym.
AGRIS	Agrindex	FAO	1975-	500 000	maatalous	artikkeleja, kokousjulkaisuja, kirjoja, väitöskirjoja, patentteja ym.
BIOSIS	Biological Abstracts & BioResearch Index	BioSciences Information Service USA	1969-	3 200 000	luonnontieteet, lääketiede	artikkeleja, kirjoja, raportteja ym.
CA-Search	Chemical Abstracts	USA	1967-	5 000 000	kemia	artikkeleja, patentteja, raportteja ym.
CAB abstracts	n. 25 tiivistelmälehtiä	The Commonwealth Agricultural Bureaux, U.K.	1972-	1 200 000	maatalous	artikkeleja, kirjoja, sarjoja, teknisiä raportteja, kokousjulk., patentteja ym.
CRIS/USDA		U.S. Department of Agriculture, USA	1971- jatkuvasti	50 000	maatalous- ja metsätiede	USA:ssa käynnissä olevat tutkimusprojektit, säilytetään 1-2 vuotta
ENERGYLINE	Energy Information Abstracts	Environment Information Center, USA	1971-	35 000	energia	artikkeleja, kokousjulk., kongressijulk., raportteja ym.
ENVIROLINE	Environment Abstracts	Environment Information Center, USA	1971-	91 000	ympäristötieteet, luonnonvarat	artikkeleja, kokousjulkaisuja, kongressijulk., raportteja ym.
Environmental Bibliography FOREST		Environmental Studies Institute, USA	1973	180 000	ympäristötieteet	aikakauslehtiartikkelit
GEOREF	5 geologista tiivistelmälehtiä	Forest Products Research Society, USA	1947- (1979)	12 000	mekaaninen metsätöteollisuus	artikkeleja, kongressijulkaisuja, patentteja ym.
NTIS		American Geological Institute, USA	1967- (1961 USA)	700 000	geologia ja geofysiikka	artikkeleja, kirjoja, kansvälörganisaatioiden julkaisuja ym.
PASCAL	Bulletin Signalétique	National Technical Information Serv. USA	1964-	900 000	ympäristö, energia, teknologia, aluesuunnittelu	tutkimuksia, raportteja, artikkelia, kongressijulkaisuja ym.
Pollution Abstracts		Centre National de la Recherche Scientifique, Ranska	1973-	3 500 000	luonnontieteet, lääketiede, tekniikka	artikkeleja, väitöskirjoja, raportteja, patentteja, kokousjulkaisuja ym.
PTS PROMT	Pollution Abstracts	Data Courier Inc., USA	1970-	80 000	ympäristötieteet, saastuminen	artikkeleja, tutkimusraportteja, kokousjulk., kirjoja, kongressijulk.
	Predicts Overview of Markets and Technology, PROMT	Predicts Inc. USA	1972-	400 000	taloustieteet, teknologia	artikkelit, selvitykset, ennusteet
Scisearch	Science Citation Index	Institute for Scientific Information USA	1974- (1965-)	7 200 000	luonnontieteet, lääketiede, tekniikka	artikkeleja, raportteja ym.
Social Scisearch	Social Science Citation Index	Institute for Scientific Information USA	1972-	1 000 000	yhteiskuntatieteet	artikkeleja, kokousjulkaisuja ym.

AGRIS antoivat 10 % kaikista relevanteista viitteistä, erikoistuneet tietokannat 30-80 % ja monitieteelliset tietokannat 10-75 %. Harveyn (1979) mukaan puolestaan saatiin jatkuva haussa CABista metsätieteiden viitteitä noin kaksinkertainen määrä AGRICOLAan verrattuna. Toisaalta samassa tutkimuksessa katsotaan käyttäjien olevan kiinnostuneempia haun saannista kuin sen tarkkuudesta. Huomattavakin irrelevanttien viitteiden määrä siedetään, jos siten saadaan uusia relevantteja viitteitä. Tässä jatkuvassa haussa oli tarkkuus keskimäärin 38-47 %.

2.2. Viitteiden päällekkäisyys

Jokin viite voidaan saada samastakin tietokannasta kahteen tai useampaan kertaan. Tämä aiheuttaa asiakkaalle turhia lisäkustannuksia. Päällekkäisyyttä esiintyy luonnollisesti myös haettaessa samaa aihetta eri tietokannoista. Huolimatta siitä, että tietokannat indeksoivat samoja julkaisuja, on viitteiden päällekkäisyys yllättävän pieni. Eri aiheiden haussa ollaan päällekkäisyyksistä saatu varsin erilaisia tuloksia.

Brooksin (1980) tutkimuksessa verrattiin AGRICOLA, AGRIS, BIOSIS ja CAB sekä SCISEARCH -tietokantoja. Viitteiden päällekkäisyyttä eri tietokantojen välillä todettiin eniten aiheissa, jotka koskivat määrättyjä kasvi- tai eläinlajeja. Näiden hakujen tuloksena oli pääasiassa aikakauslehtiartikkeleita. Esim. Corticumia koskevassa haussa BIOSISin viitteistä saatiin 43 % myös CABista ja 23 % oli myös AGRICOLAN viiteluettelossa. Suurin osa näistä oli samoja kaikissa kolmessa tietokannassa. 54 % AGRICOLAN viitteistä oli päällekkäisiä CABin kanssa.

Koko aineistosta oli todellinen viitteiden päällekkäisyys kuitenkin vain 20 % tai vähemmän, silloinkin kun tulostuksissa oli tärkeimpinä pidettyjen lehtien artikkelia.

Cittadino ym. (1977) ilmoittaa 50 % päällekkäisyyden AGRICOLAN, BIOSISin ja CABin yhden yksittäisen lajin hausta. Oyler (1977) taas ilmoittaa 26 % päällekkäisyyden AGRICOLAN ja Forestry Abstractsin (CAB) mukaan kuitenkin tulosten erilaisuus eri tietokannoissa vahvistaa käsitystä, että päällekkäisiä viitteitä ei tule moninkertaisesti usean

tietokannan haussa. Tämä tarkoittaa myös, että monen tietokannan haut ovat oikeutettuja, jos haun halutaan olevan mahdollisimman kattava. Tarvittaisiin kuitenkin lisäselvityksiä todellisen päällekkäisyyden määrittämiseksi ja sen selvittämiseksi miksi päällekkäisyyttä ei ole silloinkaan, kun tietokannat periaatteessa indeksoivat samaa aineistoa.

Viitteiden päällekkäisyys CAB-tietokannan sisällä on keskimäärin 10-15 %. Se johtuu viitteiden sijoittamisesta moniin eri osatietokantoihin. Tämän vuoksi myös saaduista viitteistä on 15-30 % ollut lisäviitteitä, joissa on epätäydelliset bibliografiset tiedot (Brooks 1980).

Cittadinon ym. (1977) AGRICOLA, BIOSIS ja CAB -vertailussa oli tietokannan sisäinen viitteiden päällekkäisyys myös suurinta CABissa, jossa 26 % viitteistä oli kahteen kertaan. AGRICOLAssa oli päällekkäisyyttä 9 % ja vähiten BIOSISissa, alle 1 %. Päällekkäisyys lisää haun kustannuksia, tässä tapauksessa eniten CABissa.

Eri tietokantojen välistä päällekkäisyyttä todettiin tässä vertailussa runsaasti. Noin puolet jonkun tietokannan viitteistä todettiin myös jossakin toisessa; suurin osa näistä jopa kaikissa kolmessa tietokannassa.

Eniten ainutkertaisia viitteitä, joita ei ollut lainkaan kahdessa muussa tietokannassa, oli CABissa, jossa oli myös kokonaismäärältään eniten viitteitä. Kaikista saaduista viitteistä oli vain CABissa 24 %, vain AGRICOLAssa 22 % ja vain BIOSISissa 13 % (v. 1973).

Longon & Machadon (1981) AGRICOLA, AGRIS ja CAB -vertailussa todettiin, että tietokantojen päällekkäisyyttä on kausijulkaisunimekkeiden perusteella tietokannoissa parittain vertailtuna 19-24 %. Kaikissa kolmessa tietokannassa on yhteistä päällekkäisyyttä vain 10 %, joka vahvistaa kirjoittajien mukaan niiden erillisyyttä tietokantoina alueellisen ja aiheittaisen kattavuuden suhteen. Harveyn (1979) mukaan metsätieteissä CABin ja AGRICOLAN päällekkäisyys on vain 2 %.

Viitteitä, jotka saatiin vain yhdestä tietokannasta (tyypillisissä metsätieteiden kysymyksissä) oli Bruggen ym. (1979) vertaamisessa 21 tietokannassa eniten CABissa, 39 %, vähemmän AGRICOLAssa, 16 %, AGRISissa 2 %, BIOSISissa 14 % ja PASCALissa 10 %. CABin Forestry Abstracts -osatiekannassa näitä oli 29 %.

2.3. Kattavuus

Eri tietokantojen kattavuuksia on vertailtu maantieteelliseltä kannalta alueittain ja toisaalta sen mukaan, mitä aihealoja ja kuinka täydellisesti ne peittävät. Maantieteelliset vertailut on tehty tietokantoihin indeksoitujen lehtien luetteloiden perusteella. Eri alojen kattavuutta tietokannoissa on selvitetty eri aiheista tehtyjen hakujen antamien tulosten perusteella.

2.3.1. Alueellinen kattavuus

Longon & Machadon (1981) tutkimus perustuu CAB, AGRIS ja AGRICOLA -tietokantojen indeksoimien kausijulkaisujen luetteloihin. Luettelot on laadittu vuosina 1973–76. Niiden perusteella saatiin näiden kolmen tietokannan alueelliseksi kattavuudeksi 153 maata. CAB indeksoi 8500 kausijulkaisunimekettä, AGRICOLA 6079 ja AGRIS 5365 nimekettä. Näistä kolmesta tietokannasta on vertailtu pareittain kahta kerrallaan ja selvitetty niiden paremmuutta eri maanosien kattavuuksissa. Pohjois-Amerikassa on AGRICOLAn ja CABin yhteinen kattavuus suurin, 90 % analysoiduista nimekkeistä. Aasiassa peittävät AGRIS ja CAB yhdessä 84 %. Euroopassa on AGRISin ja CABin peittävyys suurin 84 %, toisena AGRICOLA ja CAB 80 % ja kolmantena AGRICOLA ja AGRIS 65 % analysoiduista nimekkeistä.

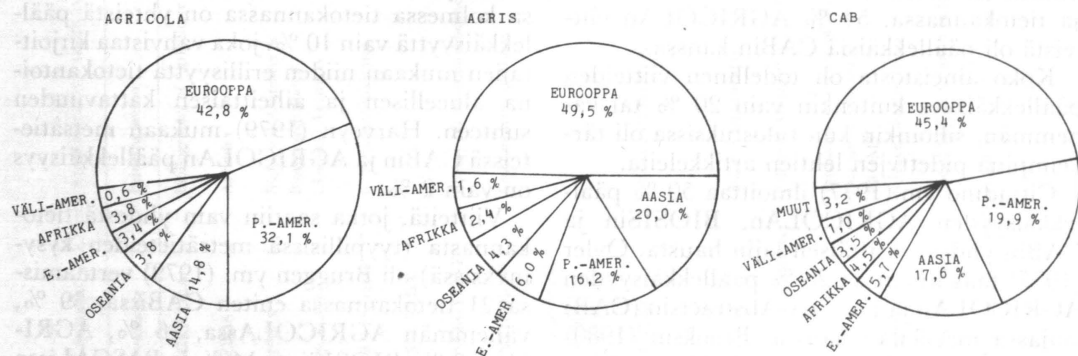
Selvityksen mukaan AGRIS kattaa sosialistiset maat huomommin kuin CAB, joka

kattaa ne näistä kolmesta tietokannasta parhaiten. AGRIS sen sijaan kattaa parhaiten Ranskan ja Japanin. Suurin osa nimekkeistä on kaikissa kolmessa tietokannassa Yhdysvalloista.

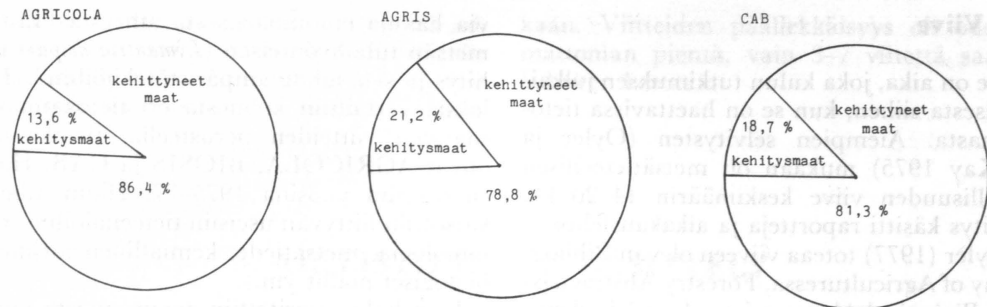
CAB kattaa parhaiten kausijulkaisuja ja alueellisesti eri maita. Sillä on vähemmän indeksoituja artikkeleita vuotta kohti ja siis vähiten artikkeleita keskimäärin kutakin julkaisua kohti vuodessa. Se on siis varsin valikoiva. AGRISissa on puolestaan vähiten indeksoituja julkaisuja, mutta kuitenkin eniten indeksoituja artikkeleita vuodessa. Se on siis vähemmän valikoiva.

Maanosista oli tietokannoissa parhaiten edustettuna Eurooppa (kuva 2): AGRIS noin 50 %, AGRICOLA 43 % ja CAB 45 %. Toiseksi parhaiten edustettuna on Pohjois-Amerikka: AGRICOLA 32 %, CAB 20 % ja AGRIS 16 %. Toiseksi parhaiten on AGRISissa edustettuna Aasia 20 %, jonka osuus AGRICOLAssa on 15 % ja CABissa 18 %, Etelä-Amerikan edustus on tietokannoissa keskimäärin 3–6 %. Kausijulkaisujen nimekkeistä on tietokannoissa suurin osa maista, joiden kansantulo on yli \$ 1 500 henkeä kohti vuodessa, eniten AGRICOLAssa 86 % (kuva 3). Kehitysmaiden osuus on 14–21 % (kuva 3).

Niistä 153 maasta, joita on tarkasteltu, eivät nämä tietokannat indeksoineet lainkaan kolmeakymmentä maata (20 %). Nämä maat ovat pääosin kehitysmaita (tulot alle \$ 1 500 henkeä kohti), niiden yhteinen pinta-ala on lähes 8 miljoonaa neliökilometriä ja väestö lähes 82 miljoonaa. Näistä 30 maasta sijaitsee kolmannes Afrikassa, 10 % Keski-Amerikas-



Kuva 2. Aikakauslehtinimekkeiden kattavuuden jakautuminen mantereittain AGRICOLA-, AGRIS- ja CAB-tietokannoissa, 1979. Longo & Machado 1981.



Kuva 3. Aikakauslehtinimekkeiden kattavuuden jakautuminen maiden kehitysasteen mukaan AGRICOLA-, AGRIS- ja CAB-tietokannoissa, 1979 (kehittyneet maat = tulot henkeä kohti yli \$ 1500 vuodessa). Longo & Machado 1981.

sa, 43 % Aasiassa ja 7 % Euroopassa. Vain 80 julkaisua indeksoidaan kaikissa kolmessa tietokannassa.

Neuvostoliiton ja muiden sosialististen maiden osuus tietokantojen aineistoista on yleensä 10 %:n paikkeilla. Bibliography of Agriculturessa, joka on AGRICOLAn papeiritulostus, neuvostoliittolaisen aineiston osuus oli Sheininan (1978) mukaan 7–15 %. Huomattava osa neuvostoliittolaisista julkaisuista on englanninkielisiä käännöksiä (23 %). Bibliography of Agriculture sisältää artikkeleita 90 neuvostoliittolaisesta aikakauslehdessä. Sarjajulkaisut ovat huonosti edustettuina.

2.3.2 Aiheittainen kattavuus

Metsätieteiden ja eri osa-alueiden sisällymistä niin painettuihin sekundaarijulkaisuihin kuin niitä vastaaviin tietokantoihinkin on pyritty selvittämään sekä perinteisin että atkmenetelmin. Elbournin (1979) selvityksen mukaan vuosittain koko CABiin lisättävistä 140 000 viitteestä liittyy metsätalouteen ja metsäntuotteisiin n. 8 000 viitettä eli 5,7 % koko aineistosta. Leatherdale (1979) puolestaan toteaa, että AGRIS-tietokannassa metsäalan viitteitä on 3–4 %. Tästä 24 % on ns. ei-konventionaalista aineistoa. Brooks (1980) tutkimuksessa vertailtiin AGRICOLA, BIOSIS, CAB ja SCISEARCH-tietokantoja. BIOSIS ei kattanut juuri lainkaan maatalous- ja luonnonvarojen ekonomian alan lehtiä, AGRICOLA oli tämän suhteen hieman kattavampi kuin CAB. BIOSIS näyt-

tää antavan hyvän tuloksen yleisekologisista sekä tauteja ja metsän tuholaisia koskevista aiheista. CAB kattaa ehkä tasaisimmin kaikki metsätieteiden osa-alueet ja se sisältää eniten maaperätutkimuksia muihin tietokantoihin verrattuna. Eräistä hyönteislajeista saatiin melko samanlaiset tulokset eri tietokannoista. Sensijaan yleinen käsite ”biologinen torjunta” tuotti kaksi kertaa enemmän viitteitä CABista kuin AGRICOLASTA.

2.4. Dokumenttien muoto

Brooksin (1980) mukaan edellä mainittujen hakujen tulokset jakautuvat dokumentin muodon mukaan siten, että viitteistä oli CABissa 55–70 % aikakauslehdistä ja muista kausijulkaisuista, 20–30 % teknisistä raporteista, 10–15 % monografioista ja 5–10 % kokous- ja konferenssijulkaisuista. AGRICOLAn viitteistä oli 80–90 % aikakauslehtiä ja muita kausijulkaisuja, 5–10 % raporttikirjallisuutta ja 5–10 % monografioita. BIOSISin viitteistä oli 80–90 % aikakauslehtiartikkeleita, 10 % monografioita ja 10 % kokousjulkaisuja. Tulokset perustuvat kuitenkin vain muutamiin hakuihin eivätkä ehkä ole edustavia.

Croninin (1980) mukaan katsauksilla on yhä tärkeämpi tehtävä tieteellisen tutkimuksen kokoamisessa ja levityksessä. Kirjallisuuskatsauksia toivottiin CABin käyttäjien taholta lisää.

2.5. Viive

Viive on aika, joka kuluu tutkimuksen julkaisemisesta siihen, kun se on haettavissa tietokannasta. Aiempien selvitysten (Oyler ja McKay 1975) mukaan oli metsätieteellisen kirjallisuuden viive keskimäärin 14–20 kk. Selvitys käsitti raportteja ja aikakauslehtiä.

Oyler (1977) toteaa viiveen olevan Bibliography of Agriculturessa, Forestry Abstractsissa ja Biological Abstractsissa yleensä korkeintaan 15 kuukautta (90 % dokumenteista). Cittadino ja muut (1977) katsovat sensijaan AGRICOLAN ja BIOSISin olevan paremmin ajan tasalla kuin CAB. CABin pitkä viive tulee todennäköisesti jatkumaankin, koska viitteiden suurta informaationvälitystä tiivistelmien pidetään nopeutta tärkeämpänä.

Brooksin (1980) vertailemissa AGRICOLA-, CAB- ja BIOSIS-tietokannoissa selvitettiin viimeisimmän lehden numeron tuleminen kuhunkin tietokantaan. Pisin viive oli CABilla. CABilla ja BIOSISilla viivettä lisää tiivistelmien teko. Brooksin mukaan tutkijat voivatkin parhaiten käyttää tietokonehakuja takautuviin hakuihin tai hakea oman alansa ulkopuolelta monitieteellisiä aiheita. Äskettäin julkaistu aikakauslehtikirjallisuus tai tutkimusraportit kannattaa usein etsiä Current Contents -julkaisusta tai aikakauslehtien sisällysluetteloista.

2.6. Hakusuunnitelma

Tietokonehaut on useimmiten tehtävä eri tavoin eri tietokannoista riippuen kunkin tietokannan rakenteesta. Brooksin (1980) tutkimuksessa haettiin aiheita ”Hyönteis- ja muut eläintuhot havupuiden kukissa ja kävyissä”. Aiheittaisia (category) ja biosystemaattisia koodeja käytettiin silloin kun se oli mahdollista. Laajat käsitteet ”havupuut” ja ”tuoheoläimet” rajoitettiin hakutermeihin ja otsakkeisiin sen varmistamiseksi, että ne olisivat melko keskeisiä aiheita saatavissa dokumenteissa. Brooksin mukaan on hyvin mahdollista, että viitteitä voitaisiin saada enemmän, jos haku tehtäisiin myös kaikkien mahdollisten tuoheoläinten nimillä ja niiden tieteellisillä nimillä. Tämä olisi kuitenkin varsin kallista.

Cittadino ym. (1977) ovat tehneet takautu-

via hakuja entomologisesta aiheesta ”Erään metsän tuhohyönteisen (*Lymantria dispar*) kehitys ja sen suhde ympäristötekijöihin”. Tuloksia vertailtiin kolmesta eri tietokannasta saatujen viitteiden perusteella, tietokannat olivat AGRICOLA, BIOSIS ja CAB. Haut suoritettiin vuosina 1975–77. Haun aiheen katsottiin liittyvän useisiin tieteenaloihin (entomologia, metsätiede, kemiallinen torjunta, biologiset mallit ym.).

Kun haku suoritettiin avainsanoista muodostetun profiilin avulla jäi esiin saamatta monia tärkeitä viitteitä. Tämän katsottiin todennäköisesti johtuvan siitä, että avainsanat eivät yksinkertaisesti esiintyneet nimekkeessä eivätkä hakusanoina.

de Jong-Hofman (1981) on vertaillut kolmea tietokantaa tekniikan alaan liittyvässä haussa (vrt. myös Langlois 1977). Tutkimuksessa selvitettiin tietyn lehden yhtä aihetta koskevat artikkelit muutaman vuoden ajalta (25 kpl) ja tutkittiin niiden löytymistä seuraavista bibliografioista: COMPENDEX, INSPEC ja PASCAL. Selvitetiin artikkelien määrä, tiivistelmien laatu ja indeksitermien sopivuus, tarkkuus ja tehokkuus ym. Kävi ilmi, että julkaisujen valinta näihin kolmeen tietokantaan oli hyvin erilaista.

Aiheille valittiin sopivat hakusanat ja näillä termeillä suoritettiin viisi eri hakuja kustakin tietokannasta. Lisäksi tehtiin käsinhaku painetuista julkaisuista. Haun tulos, tarkkuus ja päällekkäisyys näytti olevan riippuvainen valitusta hakustrategiasta. Lisäksi tietokantojen välillä todettiin suuria eroja.

Tiivistelmien tasossa oli toivomisen varaa. On huomionarvoista, että tärkeitä käsitteitä oli jäänyt pois tiivistelmästä, tässä tapauksessa esim. COMPENDEXissa 25 %, INSPECissä 29 % ja PASCALissa 41 %. Aiheeseen kuuluvia tesaurus-termejä oli hyvin ensin mainitussa, muissa vain laajempi määrittely oli mahdollinen. Otsakkeen sanat eivät vastanneet kovinkaan paljon sisältöä.

Hakumuotoja vertailtiin tekemällä laajoja ja suppeita hakuja. Sama haku tehtiin viidellä tavalla kaikista kolmesta tietokannasta. Vaikka valittu aihe on yksinkertainen, oli erittäin vaikea päätellä paras hakustrategia. Hakijalle onkin tärkeää tietää tapa, jolla julkaisijat järjestävät bibliografiat, mitä julkaisuja indeksoidaan, millaista valintaa tehdään ja millä tarkkuudella mukaan otettua kirjallisuutta indeksoidaan ja voidaan hakea.

On merkillepantavaa, että kaikkia 25 julkaisua ei saatu esille millään hakustrategialla edes kaikista kolmesta tietokannasta yhteensä. Seitsemää julkaisua ei saatu esiin lain-

kaan. Viitteiden päällekkäisyys oli odottamattoman pientä, vain 3–7 viitettä saatiin kolmeen kertaan.

3. AINEISTO JA SEN KÄSITTELY

Helsingin yliopiston metsäkirjaston tutkimusprojektissa selvitettiin muutamien kansainvälisten tietokantojen soveltuvuutta suomalaisten metsätutkijoiden tarpeisiin. Tarjoukset maksuttomista metsätieteellisistä kirjallisuushauista lähetettiin seuraavasti:

- Helsingin yliopiston maatalous-metsätieteellisen tiedekunnan laitokset 13, vastauksia saatiin yhdeksän
- Suomen Akatemian tutkijat 20, vastauksia saatiin 15
- Muita, ”spontaneja” hakupyynnöitä kolme

Taulukko 2. Hakujen aiheet ja käytetyt tietokannat.

1. Drainage norms	CAB, AGRICOLA	14. The share of stem, branch and needle growth dependent on the within-stand light regime in a young Scots pine stand	CAB, BIOSIS
2. The ideotype in forest tree breeding	CAB, AGRICOLA	15. Ecophysiology of dormancy of some boreal forest plants	CAB, BIOSIS
3. The environmental factors affecting conifer plants in natural regeneration	CAB, AGRICOLA	16. Ecology of the pine bark-bug, <i>Aradus cinnamomeus</i>	CAB, AGRICOLA, BIOSIS
4. Forest taxation (policy)	CAB, AGRICOLA	17. Formation of photosynthesis in Scots pine transplants	BIOSIS
5. Potential evapotranspiration from different vegetation types. Measurement and comparison	CAB, AGRICOLA, BIOSIS	18. Effects of air-borne pollutants on forest species and forest ecosystems	BIOSIS, ENVIROLINE
6. Stress and strain in tree seedlings between lifting and planting	CAB	19. The crystal formation in the chloroplast stroma area	CAB, BIOSIS
7. Forest succession and its mathematical models	AGRICOLA	20. Evapotranspiration of coniferous trees (in northern coniferous belt)	CAB, AGRICOLA, BIOSIS
8. The role of environmental stresses in diseases of forest trees	CAB	21. Marketing strategies and marketing channels in forest products marketing	CAB, AGRICOLA, FOREST PRODUCTS, MANAGEMENT CONTENTS
9. Effect of macronutrients and heavy metals on the physiology of mosses	AGRICOLA, BIOSIS, ENVIROLINE	22. Succession of ground vegetation (clear cutting areas)	CAB, BIOSIS
10. The world-wide use of exotic tree species	CAB	23. Subsidence of peat caused by drainage	CAB, AGRICOLA, BIOSIS
11. Biomass tables construction by regression analysis for pine and spruce trees	CAB, AGRICOLA	24. Effectiveness of forest policy	CAB, AGRICOLA
12. Effect of shading (light intensity) on acclimatization, photosynthesis, dry matter production and needle shape	BIOSIS	25. Perspectives of combined agriculture and farming	CAB, AGRICOLA
13. Production of honeydew in coniferous forests	CAB, AGRICOLA	26. Effects of nitrogen fertilization on the biological activity and nitrogen fixation in the humus layer of coniferous forest soils	CAB, AGRICOLA, BIOSIS

tötarkoitusta sekä kustannuskysymyksiä. Kyselylomakkeen palautti 27 toimeksiantajasta 26.

Haut tehtiin Decwriter LA-34 -pääteellä. Yhdysvalloissa sijaitseviin tiedostoihin oltiin

yhteydessä aluksi Amsterdamin solmun kautta. Marraskuun aikana siirryttiin Finnpack-yhteyteen Tukholman kautta. Viimeksimainnitu oli edellistä edullisempi ja toimi suhteellisen hyvin kokeilun aikana.

4. TULOKSET

4.1. Tarkkuus, päällekkäisyys ja viitteiden jakautuminen tietokannoittain

Toimeksiantajista 62 %:lle tarjottu atk-pohjainen kirjallisuushaku oli ensimmäinen. Haun käyttötarkoitukseksi ilmoitettiin seuraavaa:

		%
tutkimus	17	63,0
josta väitöskirja	8	
josta lisensiaattityö	6	
esitelmä	4	14,8
opetus	3	11,1
artikkelin kirjoittaminen	2	7,4
muu	1	3,7
	27	100,0

Haun käyttötarkoitus on hyvä tietää formulointivaiheessa haun laajuutta päätettäessä. Suurin osa toimeksiantajista käytti odotetusti tuloksia tutkimustarkoituksiin. Yksi vastaajista oli selvittänyt muuksi käyttötarkoitukseksi: "ei käyttöä". Yhden haun tulokset käytettiin sekä esitelmään että opetukseen.

Viitteitä saatiin kaikista hauista keskimäärin 102,5 hakua kohti. Tulostettujen viitteiden määrä hakua kohti vaihteli nolasta 480:een. Haut tehtiin keskimäärin 2,1 tietokannasta hakua kohden. Jotkut hauista tehtiin eri formuloinneilla samasta tietokannasta. Jos nämä otetaan huomioon eri tietokannoista tehtyjen hakujen lisäksi, saadaan yhteensä eri formulointeja tai tietokantoja 2,5 kpl hakua kohti.

Saaduista viitteistä oli vastaajien mukaan relevantteja eli vastasi annettua aihetta 53 %. Relevantteja, irrelevantteja ja päällekkäisiä viitteitä saatiin seuraavasti:

	kpl	kpl hakua kohti	%
relevantteja viitteitä	1419	54,6	53,2
irrelevantteja viitteitä	1151	44,3	43,2
päällekkäisiä viitteitä	95	3,6	3,6
yhteensä	2665	102,5	100,0

Vastaajilta tiedusteltiin myös, kuinka moni niistä viitteistä, jotka he olivat merkinneet irrelevantteiksi, oli muuten kiinnostavia. "Muuten kiinnostavia" -viitteitä oli 202 kpl eli 17,5 % irrelevantteista viitteistä.

Päällekkäisiä viitteitä eli saman viitteen esiintymistä tulostuslistalla kahteen tai useampaan kertaan ei erikseen tiedusteltu vastauslomakkeilla. Niitä oli kuitenkin havaittu yhteensä 95 kpl eli 3,6 %. Päällekkäisyyksien toteaminen on varsin työlästä suurten viitemäärien ollessa kysymyksessä. On todennäköistä, että kaikkia päällekkäisyyksiä ei ole havaittu, joten lukua on pidettävä päällekkäisyyden vähimmäisarviona.

Saadut viitteet jakautuivat tietokannoittain seuraavasti:

Tietokanta	viitteitä yht.	%	käyttö-kertoja	%
CAB	1048	39,3	21	38,9
AGRICOLA	331	12,4	16	29,6
BIOSIS	847	31,8	13	24,0
ENVIROLINE	428	16,1	2	3,7
FOREST PRODUCTS	6	0,2	1	1,9
(MAN. CONT.)	5	0,2	1	1,9
	2665	100,0		100,0

Relevantit viitteet jakautuivat tietokannoittain puolestaan seuraavasti:

Tietokanta	relevantteja viitteitä	tarkkuus %
CAB	535	51,0
AGRICOLA	193	58,3
BIOSIS	298	35,2
ENVIROLINE	387	90,4
FOREST PRODUCTS	6	(100,0)
	1419	

Eniten relevantteja viitteitä saatiin CAB-tietokannasta. AGRICOLAn tarkkuus oli varsin hyvä, samoin CABin. BIOSISin tarkkuus jäi näissä hauissa pienemmäksi. ENVIROLINEsta saatiin runsaasti relevantteja viitteitä siihen soveltuvasta muutamasta aiheesta.

Mielipiteet haun tuloksesta jakautuivat seuraavasti

		%
hyvä	6	23,1
melko hyvä	5	19,2
välttävä	11	42,3
huono	4	15,4
	26	100,0

Hakujen tuloksia voidaan pitää tyydyttävänä verrattuna aikaisempiin selvityksiin. Arviointeihin vaikuttaa paitsi haun tulos sinänsä myös vastaajien hakuun kohdistamat ennako-odotukset.

4.2 Aika- ja kustannustekijät

Päätetyöskentelyyn kului keskimäärin hakua kohden aikaa 24 minuuttia ja tietokantaa kohden noin yhdeksän minuuttia.

Kun hakuihin käytettiin Amsterdamin solmua, puhelinkustannukset nousivat verraten korkeiksi. Myöhemmin, siirryttäessä Finnpackin käyttöön, tietoliikennekustannukset alenivat.

Neuvotteluihin ja hakujen formulointiin kulunutta työaikaa, laitemenoja ja muita "juoksevia menoja" ei ole otettu huomioon kokonaiskustannuksia laskettaessa.

Kustannusten muodostuminen hakua kohden oli seuraava:

		%
tietokoneaika	71 mk	27,3
tulostetut viitteet	68 mk	26,2
tietoliikennekulut (arvio)	121 mk	46,5
Yhteensä	260 mk	100,0

Vastaajia pyydettiin arvioimaan haun tulos ja tiedusteltiin heidän halukkuuttaan käyttää tietokonepohjaisia kirjallisuushakuja sittenkin kun palvelusta tulee maksullinen. Mielipidettä tiedusteltiin myös tavallisen keskihintaisen haun kustannusten kohtuullisuudesta.

Kysymykseen "Ovatko käyttäjät halukkaita käyttämään atk-palveluja, kun palvelu tulee maksulliseksi?" saatiin seuraavat vastaukset:

		%
kyllä	23	88,5
ehkä	2	7,7
ei	1	3,8
yht.	26	100,0

Käyttäjät näyttävät olleen hakutuloksiinsa jossain määrin pettyneitä. He ovat kuitenkin halukkaita käyttämään atk-palvelua kirjallisuuden haussa sittenkin, kun palvelu tulee maksulliseksi. Hakujen tulosten uskotaan ehkä paranevan järjestelmien kehittymisen ja käyttökokemuksen myötä. Ilmeisesti vastaajat otaksuvat tällä menetelmällä saavansa tietoa, jota ei muilla keinoilla saataisi tai jota ei saataisi yhtä vaivattomasti tai edullisesti.

Kysymykseen pitävätkö käyttäjät 200-300 markan kustannusta yhdestä tietokonehausta kohtuullisena vastattiin:

		%
kyllä	18	69,2
ehkä	5	19,2
ei	3	11,6
yht.	26	100,0

Kustannusten kohtuullisuudesta olivat vastaajat tehneet joitakin huomautuksia. Kohtuullisuuden katsottiin riippuvan ta-

pauksesta. Yksi tutkija olisi valmis maksamaan uuden, laajan haun onnistuneesta tuloksesta 1000 markan tutustumisrahan, mutta välttävistä ja huonosta tai pikkuhausta ei mitään tai muutaman kympin. Toisessa vastauksessa puolestaan arveltiin kohtuulliseksi hinnaksi 100–200 markkaa. Tämän vastaajan haussa oli viitteitä 157, joista relevanteiksi oli merkitty 92.

5. TULOSTEN TARKASTELUA

Käytetyin tietokanta oli CAB, jota käytettiin 21 haussa (38 %) (taulukko 2). Kaikista tulostetuista viitteistä 39 % oli peräisin CABista, tarkkuus oli 51 %. CABin käyttö on edullisempaa kuin AGRICOLAn ja BIOSISin käyttö, koska tietokantaa ei ole jaettu, vaan ajanjakson 1972–81 aineisto voidaan hakea yhdellä kertaa. CABin viitteiden hinnat ovat edellisiä tietokantoja kalliimpia, mutta sisältävät hyvät tiivistelmät.

Toiseksi eniten käytettiin AGRICOLA-tietokantaa, josta haettiin kirjallisuusviitteitä 16 aiheesta. Näistä 14 hakua tehtiin myös CABista, 6 BIOSISista ja yksi pelkästään AGRICOLAsta. Kaikista käytetyistä tietokannoista AGRICOLAn osuus oli 29 %. Siitä saatiin vain 12 % kaikista tulostetuista viitteistä, mutta tarkkuus oli keskimääräistä parempi eli 58 %.

BIOSIS-tietokannasta tehtiin haut 13 kysymykseen, joista kahteen pelkästään siitä. Yhdeksään kysymykseen haettiin kirjallisuutta myös CABista ja viiteen myös AGRICOLAsta. Kaksi hakua tehtiin BIOSISin lisäksi myös ENVIROLINEsta.

BIOSISin osuus käytetyistä tietokannoista oli 23 % ja saaduista viitteistä 32 %. Tarkkuus oli keskiarvoa alhaisempi eli 35 %. BIOSISin indeksointi- ja luokitusjärjestelmä poikkeaa paljon CAB ja AGRICOLA-tietokannoista, joten hakusuunnitelma on tehtävä nimenomaan tälle tietokannalle erikseen. Biologisissa aiheissa oli relevanttien viitteiden määrä kuitenkin sama tai suurempi kuin CABista saatujen ja huomattavasti suurempi kuin AGRICOLAsta saatujen määrä.

ENVIROLINE-tietokantaa on tässä selvityksessä voitu käyttää vain kokeilunomaisesti kahdessa haussa. Aihepiirinsä vuoksi useim-

Kustannuksia kohtuullisina pitäneistä vastaajista eräs mainitsee kirjallisuudenhaun perinteisin menetelmin omasta tutkimusaiheestaan olevan osittain mahdollonta sopivien lutteloiden puutteen vuoksi. Kahdessa vastauksessa todetaan, että hinnan kohtuullisuus riippuu tutkimustilanteesta ja määrärahoista, jotka eivät moniin kirjallisuushakuihin riitä.

mat metsäalan haut eivät soveltuneet siihen. On kuitenkin ilmeistä, että niistä aiheista, jotka liittyvät ympäristönsuojeluun tai -vaurioihin, voidaan ENVIROLINEsta saada huomattavasti parempi tulos kuin muista tietokannoista. Pienestä Forest (Products)-tietokannasta tehtiin vain yksi haku, joten sen käyttökelpoisuudesta ei voida tehdä johtopäätöksiä. CAB-tietokannan sisältämä Forestry Abstracts osatietokanta sisältää noin 60.000 viitettä koko CABin 1.400.000 viitteestä. Forestry Abstracts on metsätieteellinen tietokanta, jota tässä selvityksessä ei ole erotettu koko CABista. Joissakin tapauksissa Forestry Abstractsin käyttö erikseen on ilmeisesti hyödyllistä, vaikka se onkin suppea.

Tässä selvityksessä käytetyt tietokannat on rakennettu eri tavoin. Aineisto kussakin tietokannassa on erilaista. Se on syötetty, indeksoitu ja luokitettu eri periaatteita ja järjestelmiä noudattaen. Eräinä tärkeimpinä puutteina metsätieteellisen informaation välityksen kannalta voidaan pitää seuraavia seikkoja:

- kansainvälisesti hyväksytyin, sisällön analyysissa ja jälleen haussa käyttökelpoisen tesaauruksen eli hierarkisesti järjestetyn sanaston puuttuminen
- kieliongelmat yleensä, pienten maiden ja kielialueiden ongelmat
- tietojärjestelmien rakenteen erilaisuus

Mm. IUFROn (Working Parties 56.03.02/03 Terminology/Information Retrieval) ja FAOn toimesta on saatu joitakin tuloksia edellä mainittujen puutteiden poistamiseksi tai vähentämiseksi. Yksittäisen käyttäjän kannalta merkittäviä näkökohtia haun onnistumiseksi ovat mm. seuraavat:

- Haun aiheen määrittäminen hakupyynnön yhteydessä. Aiheen kirjallisessa kuvailussa tulisi olla perusidea kiteytettynä. Mukana saisi olla myös englanninkielisiä avainsanoja.
- Hakua voidaan rajoittaa ajallisesti. Tietokannat alkavat 1970-luvun alkuvuosista, vanhempiä viitteitä niistä ei yleensä ole mahdollista saada. Haku voidaan haluttaessa rajoittaa vain viimeisiin vuosiin tai muihin määrättyihin vuosiin. Toisaalta tiedostoissa ei ole aivan uusintakaan aineistoa syöttöön kuluvan viiveen vuoksi. Eri tietokannoissa on eripituinen viive, esim. CABissa se on varsin pitkä tiivistelmien laatimisen vuoksi.
- Haku voidaan rajoittaa julkaisun kielen perusteella. Maatalouskirjaston AGRIS-kokeilussa oli viitteitä pidetty arvottomina kielen takia (Ylinen & Kaivosoja 1981), vaikka ne aiheensa puolesta olivat relevantteja. Otaksutaan, että ainakin tutkijat olisivat kiinnostuneita "harvinaisillakin" kielillä julkaistusta materiaalista. Näin erityisesti siksi, että esim. CAB-tietokannassa on tiivistelmät kaikenkielisestä materiaalista. Tämä saattaa usein olla ainoa keino saada julkaisusta englanninkielinen tiivistelmä.
- Haun käyttötarkoitus tulisi tietää sen vuoksi, että se voi vaikuttaa haun laajuuteen. Perustutkimusta varten tehdään ehkä laaja haku, jolloin viitteistä saattaa olla vain pienhkö osa relevantteja. Laaja haku varmistaa kuitenkin, ettei mitään jää pois. Jos haun tarkoitus on yleisluontoisempi, esim. esitelmä tai opetus, saattaa suppeampi, tarkka haku riittää, jolloin saadaan muutamia keskeisiä viitteitä.
- Mitä "yksinkertaisempi" hakupyynnö on eli mitä vähemmän osatekijöitä se sisältää sitä paremmin se soveltuu tietokonehauksi. Selkeä ytimekäs aihe, esim. lajinimi voi tuottaa runsaasti hyviä viitteitä. Ekologistyyppiset aiheet, jotka ovat monimutkaisia, monia syy-seuraussuhteita sisältäviä, voivat olla vaikeita tai ainakin kalliita hakea. Toisaalta tällaiset haut voisivat olla käsin haettuna vieläkin hankalampia. Tietokone-

pohjainen tiedon ja kirjallisuudenhaku on yksi muiden hakumenetelmien joukossa. Se sopii joihinkin aiheisiin, joihinkin ei lainkaan.

- Atk-hausta saattaa esim. tutkijalla olla epärealistisia odotuksia. Tietokonetta käytettäessä ollaan muilla aloilla totuttu tarkkoihin tuloksiin. Useimmille kyselyyn vastanneille atk-kirjallisuushaku oli ensimmäinen. Odotukset eivät ehkä olleet realistisia, koska tietoa tietokonehaun mahdollisuuksista ja rajoituksista on puuttunut. Jotkut melko pitkällekin aihettaan tutkineet tutkijat eivät itse ilmoittaneet tuntevansa aiheestaan kovinkaan montaa viitettä. Jos hausta saatiin heille muutamia uusia viitteitä, tuntuisi tulosinformaatikon kannalta tyydyttävältä.

Asiakkaan kannalta huonosti onnistuneen haun ennakkoinformaatioissa on saattanut tapahtua virhe. Asiakkaan oma muotoilu aiheesta saattaa olla väärä tai informaation tulkinta kysymyksestä väärä. Tai lopulta tiedostojen käyttämä indeksointi on erilainen kuin hakijoiden käsitys. Hakija ei voi myöskään saada hyvää tulosta, jos hakupyynnö ei vastaa tarpeeksi hyvin kysyjän informaatiotarvetta. Tilanne on huonoin, jos hakupyynnön muoto on rajatumpi kuin todellinen tarvittava tieto.

Edellä mainittujen syiden vuoksi korostuu asiakkaan ja informaation välisen yhteistyön tärkeys. Asiakkaan oman tieteenalan terminologian tuntemus on tärkeä neuvoteltaessa haun muotoilusta, hakutermien synonyymeistä ja termien tärkeysjärjestyksestä ym. Hakustrategian laatiminen viekin paljon aikaa niin asiakkaalta kuin haun tekijältä. Haun huolellinen valmistelu on kuitenkin välttämätön hyvän lopputuloksen saavuttamiseksi.

KIRJALLISUUTTA

- BLATCHFORD, O. N. 1979. The value of ODC for information retrieval now and tomorrow. In: Schrader, S. (ed.) 1979. Proceedings of the international symposium of IUFRO Subject group S 6.03. Information systems and terminology. Mitteilungen der Bundesforschungsanstalt für Forst- und Holzwirtschaft Hamburg-Reinbek 127: 1–191.
- BROOKS, K. 1980. A comparison of the coverage of agricultural and forestry literature on Agricola, Biosis, CAB and Scisearch. Database 1: 38–49.

- BRUGGE, T., BURG, J. van der & MOLSTER, H. C. 1979. Usefulness of various computer-readable bibliographic files for the Dorschamp research institute for forestry and landscape planning. In: Schrader, S. (ed.) 1979. Proceedings of the international symposium of IUFRO . . . : 37–45.
- CITTADINO, M. L., GIESE, R. L. & CASWELL, J. V. 1977. Three computerbased bibliographic retrieval systems for scientific literature. Bioscience 11: 739–741.
- CRONIN, B. 1980. CAB abstracts: a global view. Aslib

- proceedings 11/12: 425-437.
- ELBOURN, C. A. 1979. CAB on-line services in relation to forestry. In: Schrader, S. (ed.). Proceedings of the international symposium of IUFRO . . . : 55-60.
- HARVEY, S. 1979. CAB/CAIN evaluation report: a comparative study on the performance of two agricultural databases in a computerized current awareness service. PUDOC/BLR&DD Report 5483: 1-77.
- HENRY, W. M., LEIGH, J. A., TEDD, L. A. & WILLIAMS, P. W. 1980. Online searching, an introduction. Butterworths. 209 p. London.
- de JONG-HOFMAN, M. W. 1981. Comparison of selecting, abstracting and indexing by Compendex, Inspec and Pascal and the impact of this on manual and automated retrieval of information. Online review 1: 25-36.
- HYVÄKKÄ, L. 1978. Metsätieteellinen tutkimus ja metsäkirjaston informaatiopalvelun kehittäminen. Suomen kirjallisuuspäivien X informaatiopalvelun kurssi. Erikoistyö. 16 s. Helsinki.
- LAITINEN, S. 1981. Kansainvälinen tieto tehokkaasti hallinnon käyttöön. Hallinto 7: 18-25.
- LANCASTER, F. W. 1978. Guidelines for the evaluation of information systems and services. Unisist. Unesco. 169 p. Paris.
- 1979. Information retrieval systems, characteristics, testing and evaluation. John Wiley & Sons. 381 p. New York.
- LANGLOIS, M. C. 1977. On-line information systems. Comparison of Compendex, Inspec and Ntis search files via LMS/Dialog, SDC/Orbit and SDS/Recon. Online review 3: 231-237.
- LEATHERDALE, D. 1979. AGRIS in relation to forestry. In: Schrader, S. (ed.). Proceedings of the international symposium of IUFRO . . . : 21-37.
- LONGO, R. M. J. & MACHADO, U. D. 1981. Characterization of databases in the agricultural sciences. Journal of the American society for information science 2: 83-91.
- OYLER, D. K. 1977. Indexes to forestry literature - a bibliographic appraisal. Journal of forestry 1: 20-22.
- & Mc KAY, M. W. 1975. The impact of the forest and forest inventory on the environment: a study of bibliographic coverage. University of Wisconsin-Madison. Steenbock memorial library. Bibliographic series 4.
- SCHRADER, S. (ed.) 1979. Proceedings of the international symposium of IUFRO Subject group S 6.03 Information systems and terminology. Mitteilungen der Bundesforschungsanstalt für Forst- und Holzwirtschaft Hamburg-Reinbek 127: 1-191.
- SHEININA, T. I. 1978. The Bibliography of agriculture as an information source for world agricultural literature. IAALD Quarterly bulletin 3-4: 46-48.
- SUCKCHAROEN, S. 1979. Metsäkirjaston ja sen tiedonlähteiden käyttötutkimus sekä metsäkirjaston käyttöopas. Suomen kirjallisuuspäivien X informaatiopalvelun kurssi. Erikoistyö. 37 s. Helsinki.
- YLINEN, A. & KAIVOSOJA, A. 1981. Agris-informaatiojärjestelmän on-line-kokeilu 1979-1980. Helsingin yliopiston maatalouskirjaston julkaisu- ja 3. 22 s. Helsinki.

SUMMARY

EXPERIENCES OF SOME BIBLIOGRAPHIC DATABASES IN FORESTRY

The usefulness of some international bibliographic databases for Finnish forest researchers was studied. In the study 26 gratis searches from one or more databases were made. The search results were evaluated with the aid of a questionnaire.

For 62 % of the customers the search was their first. The searches were carried out using mainly AGRICOLA, BIOSIS and CAB databases. The most used database was CAB, which includes Forestry Abstracts. The precision in CAB was 51,0 %. The number of references retrieved was lower in AGRICOLA, but the precision was slightly better, 58,3 %. The overall precision in BIOSIS, 35,2 %, was below the average, the result was better, however, in

biological or ecological topics. The costs per search were about US \$ 47 of which telecommunication costs took almost a half. AGRICOLA was the least expensive of the three databases.

The overlap between databases was rather small, this suggests that multi-database searching is necessary. About half of the users rated the results of the search good or fairly good, all were willing to use the on-line services even when they carry a charge. The opinions as to an acceptable charge for a search varied. Some previous comparable studies are reviewed and some main problems in obtaining information in forestry are discussed.