

Den metod som använts i studiet av olika personer har utvecklats av Hannes Malmberg, i diskussioner med undertecknad. Målet är att snabbt kunna sammanställa en "citerings-kurva" över ett antal personer.

## Hannes Malmberg – Metod för undersökning av enskilda i ISI

Problemen med att få en korrekt citeringsbedömning är många, och att få en fullkomligt tillförlitlig siffra är svårt. En förfrågan hos "Thomson Institute" om en metod för att göra de undersökningar som här gjorts ledde till ett svar som bedömdes som otillfredsställande, med en allt för enkel metod som skulle ha givit stora felkällor. Här har därför en betydligt mera komplicerad metod utvecklats för att få ett värde med så små fel som möjligt utan att varje post i ISI särskilt granskas.

### A. Problemen

Dubletter: Då ISI bara har skrivit in efternamn plus förnamnsinitialer finns det en stor risk att man kan få citeringarna för två, istället för en, forskare. Därför har bara forskare med namn som är unika (eller nästan unika) valts ut. En viss dubbelregistrering är svår att undvika eftersom det kan förekomma någon mindre känd vars fåtaliga verk det varit för komplicerat att sortera ut.

ISI ger en begränsad kapacitet när det gäller markering av antal verk. Problemet ligger i att när man trycker på "Select All", för en forskares alla citerade verk, så markeras bara de 500 första. Vissa forskare kan ha mer än femhundra verk citerade för ett enda år. Detta beror inte bara på deras produktivitet, utan också på att de kan finnas översatta till olika språk. Dessutom innehåller citeringsindexet olika stavningar av samma verk, antingen därför att den som citerat skrivit fel, eller eventuellt på grund av att registreringen har varit behäftad med en osäkerhet.

Det finns också en begränsad kapacitet när det gäller antal citeringar. För respektive år klarar ISI maximalt 2000 citeringar.

Av okänd anledning så sker i sorteringen en överlappning i år (en delförklaring är att vissa tidskrifter kommer ut med viss fördröjning). Exempelvis vid en sökning i tidsspannet 1995-1997 så fördelar sig citeringarna oftast mellan åren 1993-1998. Störst är överrepresentationen genom citeringar som ligger ett år tillbaka i tiden, dvs. 1994 och 1995 är betydligt större än 1993 och 1998 i en sökning som sträcker sig från 1993-1998.

Dessa problem skulle motverkas utan att undersökningen tog allt för lång tid, dvs. utan att varje enskilt verk eller citat behövde granskas särskilt. Arbetsmetoden skedde i några steg, som beskrivs detaljerat nedan.

### B. Metoden

I. Metodens första del består av urval och kontroller.

1/ I *Web of Science*, som nås från *Web of Knowledge*, väljs en forskare och ett lämpligt tidsspän. Detta är kortare om forskaren förväntas ha många citeringar. Som standard har dock fyraårsperioder använts, som 1986-1990.

2/ Sökning görs alltid på efternamnet, och begynnelsebokstaven på det första förnamnet följt av en asterisk. Exempelvis en sökning på Hobsbawm E\* inkluderar även verk skrivna av Hobsbawm E J.

3/ Så som registret är upplagt måste man efter att ha kommit in på listan över citerade verk för forskaren ifråga först trycka "Select All" och sedan "Finish Search". Då kommer ett antal sidor med citerade verk.

4/ Då varje sida har 20 verk, så kommer alla verk med om antalet sidor är <25 (dvs. färre än 500 verk). Om däremot antalet sidor är >25 så måste sökningen begränsas. Första steget är att gå tillbaka till grundmenyn och minska tidsspannet med ett eller flera år. Om det, trots att bara ett år studeras, blir mer än 25 sidor får tidsspannet när verken publicerats begränsas. Denna begränsning görs på samma sida som där namnet skrivs in. Sökningen kan då delas då in i två delar, där den ena sökningen t.ex. kan söka på verk publicerade 1940-1969 och den andra 1970-2005.

5/ En särskild kontroll görs igenom att gå genom den första sidan för att se att det inte finns några uppenbara exempel på att forskaren har en kollega med samma namn. Detta görs lämpligen genom att man försäkrar sig om att alla verk ligger någorlunda inom ämnesområdet som forskaren under granskning ägnar sig åt. Om ytterligare en forskare med samma namn förekommer måste den forskare man ursprungligen sökte efter uteslutas från undersökningen.

6/ Ytterligare en kontroll med samma syfte kan göras i sidan med listan över artiklar som *har citerat* forskaren ifråga. Här måste man dock vara uppmärksam på att forskaren ifråga kan förekomma i artiklarna som är mera disparata än forskarens egna verk. Här kan också kontrolleras att det inte finns mer än 2000 citeringar per år. Om så är fallet måste man gå tillbaka till startsidan och begränsa tidsspannet med ett eller ett par år, beroende på hur mycket mer än 2000 citeringar forskaren i fråga har.

7/ När all granskning är klar, skall man i den version som vi haft tillgång till trycka på "Analyze".

II. Därmed går man över till metodens andra del, som är resultatgenererande.

8/ På "Analyze"-sidan ska "Publication year" markeras i "Select field to rank by". Där måste man först ändra "First (up to 500 records)" till "All (up to 2000 records)". Man måste också ändra hur många resultat som visas till 25 om fler än 10 år har markerats som tidsspänn i första menyn.

9/ Därefter skall "Record Count" ändras till "Selected Field" så att åren kommer i kronologisk ordning.

10/ När listan på åren kommer upp så sträcker den sig, som tidigare nämnts, över fler år än de som markerats i första menyn. Om tidsspännet som markerades i början var t.ex. 1995-1996 är det *mycket viktigt* att 1994-resultatet tas med i räkning. Detta eftersom överskvalpningarna från 1996 till 1995 automatiskt räknas in i 1995. Om man inte räknar med 1994 och motsvarande år så kommer det konsekvent på varenda forskare att bli en oerättigad nedgång i slutet på varje markerat tidsspänn. (Jämför problem nr 4 ovan.)

De citeringar som ligger två år bak i tiden eller ett år fram är allt som oftast marginella och kan därmed exkluderas.

III. Slutligen kommer metodens tredje del, som är den statistiska bearbetningen.

11/ Klistra in de resultat som ska användas i Excel, och fortsätt med samma procedur tills alla år har varit med i en sökning. Denna granskning avslutas med att addera alla årtal det finns två av.