



DEN UMULIGE INFORMATIONSSØGNING – OM TRE DANSKE SØGESTEDERS BEGRÆNSNINGER

I vores arbejde som informationsspecialister ved Roskilde Universitetsbibliotek forsøger vi dagligt at overbevise studerende og forskere om den videnskabelige vigtighed af en systematiseret tilgang til litteratursøgning. Vi vejleder dem om brugen af booleske operatoren, om nødvendigheden af at søge på tværs af databaser og om relevansen af referencehåndteringsprogrammer.

DANIEL HONORÉ PRYN PRYN@RUC.DK, ROSKILDE UNIVERSITETSBIBLIOTEK
MIKKEL HVIDTFELDT ANDERSEN HVMA@ITU.DK, IT-UNIVERSITETET I KØBENHAVN

I februar 2015 indgik vi i forbindelse med VIOL-projektet i et samarbejde om at udforme en bibliografi over danske publikationer om velfærdsteknologi. Grundlaget for bibliografien var blandt andet en systematiseret afsøgning af henholdsvis Infomedia.dk, Bibliotek.dk og Forskningsdatabasen.dk med henblik på at sammenfatte resultaterne i en dynamisk database i referencehåndteringsprogrammet Zotero. Det var med en vis forundring, vi måtte erkende, at det tilsyneladende ikke er muligt at foretage tilfredsstillende systematiserede afsøgninger af dansk litteratur. De øgede krav til dokumentation og videnskabelig redelighed i den akademiske verden taget i betragtning er dette en trist erkendelse. Især fordi det er noget, vi bliver nødt til at tage højde for i vores fremtidige formidling til både forskere og studerende. I denne artikel forsøger vi at opridse, hvorfor vi ikke fandt det muligt, samt at redegøre for vores fremtidige ønsker til de tre webgrænseflader.

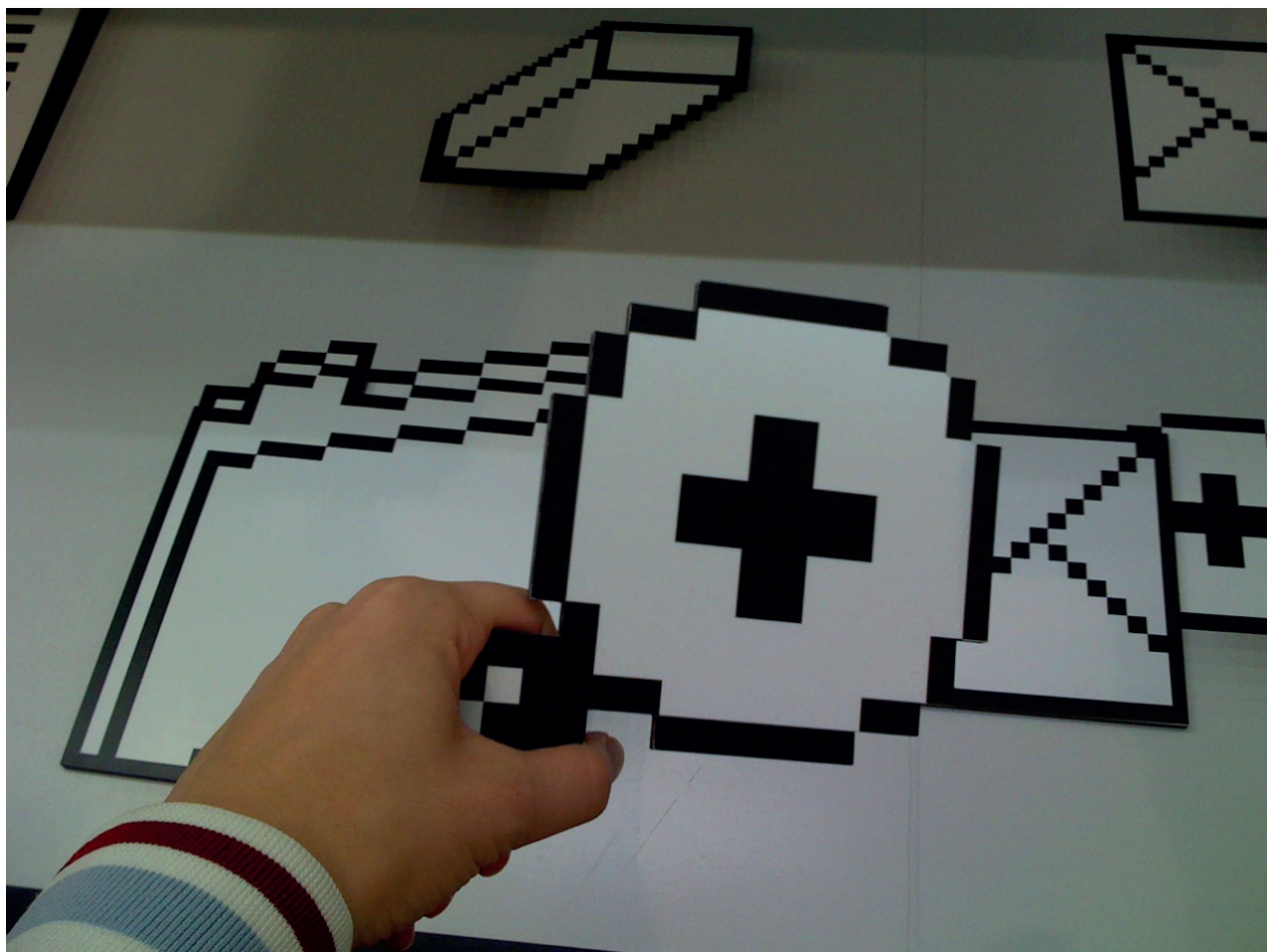
Infomedia

I VIOL-projektet blev det besluttet, at artikler fra fagblade var en væsentlig kilde at få repræsenteret i bibliografien. Infomedia var derfor det første sted, vi valgte at søge. Artiklerne skulle indeholde ordene "velfærd", "effektiv" eller "service" i kombination med enten "teknologi" eller "digital", eller blot ordet "velfærdsteknologi". De fem første søgetermer (velfærd, effektiv, service, teknologi og digital) blev trunkeret både foran og bagved, mens det sidste (velfærdsteknologi) blev trunkeret bagved. Alle søgninger var afgrænset til perioden fra og med 2008 til og med 2014 og blev foretaget i fagligt udvalgte publikationer. Selvom Infomedia muliggør fuldttekstsøgning, besluttede vi at begrænse søgningen til i titel og abstract for at sikre ensartet fremgangsmåder på tværs af databaserne.

Første søgning blev foretaget i Infomedias Ekspertsøgning, hvor vi fandt god dokumentation for søgesyntaks under fanen Hjælp. Selve søgedelen fungerede smertefrit i Infomedia, men problemet opstod, da vi efterfølgende skulle trække resultaterne ud af Infomedia og ind i referencehåndteringsprogrammet. Infomedia understøtter nemlig ikke nogen form for eksport af metadata, og på de tidsskrifter, vi søgte, var der heller ikke noget link til originalartiklen. For at få fat i denne metadata, var det derfor nødvendigt efterfølgende at benytte Google (eller tilsvarende søgemaskine) og her søge sig frem til den pågældende artikel. Når den pågældende artikel herefter var fundet på tidsskriftets egen hjemmeside, var det så muligt via et browser-plugin fra referencehåndteringsprogrammet at hente metadata. Metadata fra disse hjemmesider var dog i alle tilfælde mangelfuld, hvilket igen resulterede i, at (store dele af) metadata fra hver enkelt artikel, måtte indtastes manuelt. Samlet medførte den manglende understøttelse af referencehåndteringsprogrammer et meget stort manuelt arbejde med den efterfølgende indsamling af metadata.

Bibliotek.dk

De tidsskrifter, vi ikke havde mulighed for at finde i Infomedia, søgte vi på i Bibliotek.dk. Under sidens hjælpefunktion er der en smule vejledning at finde til Kommandosøgning. Denne vejledning er dog minimal og dækker slet ikke alle de kommandoer, der reelt er tilgængelige. Dette er velsagtens gjort for ikke at forvirre den almindelige bruger unødigt, men som professionel kunne man godt have ønsket sig et link til yderligere vejledning for avanceret brug. Med en smule Google-fu fandt vi i stedet frem til en mere avanceret vejledning på DanBibs hjemmeside. Denne vejledning er dog langt fra letlæselig – selv for



professionelle – men det lykkedes os at finde frem til de søgekoder, vi skulle bruge, og arbejdet kunne gå i gang.

Det første problem opstod, da vi forsøgte at arbejde med foranstillet trunkering. I princippet fungerer denne funktion efter hensigten, og søger man eksempelvis på *"*velfærd"*, dukker der både resultater som *"børnevelfærd"* og *"velfærdsteknologi"* op. Problemet opstår, når man sætter foranstillet trunkering ind i længere og mere avancerede søgestreng. Forsøges dette, får man enten en fejlmeddelelse om, der ikke kan søges på det, man har skrevet, eller også sker der simpelthen en timeout på siden. Dette betyder reelt, at der er søgninger, som ikke kan foretages på Bibliotek.dk, og for vores vedkommende betød det, at vi måtte gå på kompromis med vores systematik og nøjes med bagvedstillede trunkeringer.

En anden begrænsning ligger i mulighederne for at søge på årstal. Bibliotek.dk giver mulighed for at filtrere *"År lig med"*, *"År før"* eller *"År efter"* – men kun én af nævnte ad gangen. Da vi i vores tilfælde søgte litteratur udgivet mellem 2008 og 2014, var dette filter ubrugeligt. Heldigvis er det i søgestrengen

muligt at indsætte årstal enkeltvis, og da vi kun havde behov for at søge i en syvårig periode, kunne vi løse vores problem ved at afslutte søgestrengen med:

and (år=2008 or år=2009 or år=2010 or år=2011 or år=2012 or år=2013 or år=2014)

Havde vores krav i stedet været, at litteraturen skulle have været udgivet mellem 1847 og 1985, havde vi dog stået i en helt anden kattepine.

Det sidste problem med Bibliotek.dk opstod, da vi skulle trække referencerne ud fra vores søgeresultater. Bibliotek.dk understøtter for så vidt udtræk til referencehåndteringsprogrammer, og det bør her nævnes, at det metadata, der gives, er særdeles tilfredsstillende. Problemet er blot, at der kun er mulighed for at trække referencer ud enkeltvis (alternativt er det muligt at tilføje resultater til en såkaldt huskeliste og herefter trække dem ud samlet, men hver enkelt resultat skal stadigvæk tilføjes til huskelisten individuelt, så arbejdsbyrden bliver derved den samme). Ingen af de større referencehåndteringsværktøjers ellers udmærkede browser-plugins kunne heller løse opgaven i Bibliotek.dk.

Forskningsdatabasen

Som en del af den systematiske afsøgning besøgte vi Forskningsdatabasen for at foretage en klassisk emnesøgning efter samme opskrift som ovenfor. Dette viste sig dog hurtigt at være problematisk på flere måder.

Forskningsdatabasen anvender i skrivende stund deres helt egen søgesyntaks. Der anvendes ”spørgsmålsteget” (?) for trunkering, ”lighedsteget” (=) for ELLER, og ”melletrum” for OG. Det hele er dog fint dokumenteret under søgetips og var derfor ganske simpelt at gå til, omend det virker mærkværdigt, at valget ikke er faldet på en mere standardiseret syntaks. OG-funktionen, dvs. mellemrumsteget, forvoldte dog en del problemer, når vi arbejdede med længere sammensatte søgestrengene, og i praksis oplevede vi desuden, at OG i nogle tilfælde fungerede som ELLER. Der var således en del manuelt arbejde med at rydde op i posterne efterfølgende.

Anderledes udfordrende forholdt det sig med det, Forskningsdatabasen benævner søgekoder, og som muliggør søgning i specifikke metadatafelter. På trods af, at Forskningsdatabasen i en lang række tilfælde har trukket abstract-metadata ind i webgrænsefladen, er det ikke muligt at lave en søgning kun i dette felt. Derfor må man vælge mellem udelukkende at søge i titel-feltet (med kommandoen TI:) eller blot at lave en fritekstsøgning. For den gennemsnitlige informationssøger er dette formentlig ikke et særlig stort problem, måske fordi man er lykkeligt uvidende. Foretager man derimod en systematiseret afsøgning, står man i et dilemma, hvor man enten må berede sig på at få en masse ”støj” i sin søgning, eller fæstne sin lid til at værket titel er tilstrækkelig beskrivende til at blive indfanget i søgningen.

En lignende udfordring med søgekoder opstod i arbejdet med årstalsafgrænsningen. Her gav den fremgangsmåde, som Forskningsdatabasens hjælpefunktion beskriver, konsekvent færre resultater, end hvis man efterfølgende filtrerede søgningen på udgivelsesår. Også sprogekoden i dokumentationen er unøjagtigt dokumenteret. Her henvises der til tre-bogstavssprogkoderne fra ISO. I praksis viste det sig dog, at kun to-bogstavssprogkoden (LA:da) gav poster i søgningen. I det igangværende arbejde med at udvikle Forskningsdatabasen kan man derfor kun håbe på, at kommandosøgning og dokumentation udbygges og korrigeres efter gældende standarder.

Forskningsdatabasen excellerede dog i muligheden for at eksportere søgeresultaterne til referencehåndteringsprogrammer. Forskningsdatabasen tillader dels et samlet udtræk af alle søgeresultater på én gang og giver derudover fyldestgørende referenceposter indeholdende al den tilgængelige metadata.

Konklusion

På baggrund af vores erfaringer med bibliotekets forskningsunderstøttelse af VIOL-projektet opridses vi tre ønsker til fremtidens informationssøgningsgrænseflader.

1. Det er vores oplevelse at referencehåndteringsprogrammer (RefWorks, Zotero, Mendeley etc.) i stigende grad bliver taget i brug af både førsteårsstuderende og forskere. Forklaringerne på dette er formentlig mange, men konsekvensen er en stigende forventning til, at programmerne kan tilføje et element af systematik og gennemsøkelighed til både studiearbejdet og forskningen. Efterspørgslen på funktionelle referencehåndteringsprogrammer medfører et implicit krav til informationsudbydere om gode muligheder for at trække data ud til videre brug. Som vist ovenfor er det kun Forskningsdatabasen, der imødekommer dette behov nogenlunde tilfredsstillende. Det er vores håb, at man i fremtiden bestræber sig på at tage højde for gode output-muligheder i designet af systemernes webgrænseflader.

2. Systematik og gennemsigtighed er altafgørende for den videnskabelige metode. Det er derfor bydende nødvendigt, at bibliotekerne i deres bestræbelser på at understøtte forskere og studerende også kan kvalificere deres services som gennemsigtige og pålidelige. På baggrund af de erfaringer, vi har gjort os med ovenstående informationsressourcer, ser vi store udfordringer, når der kommer til muligheden for at kunne redegøre for søgningers og resultaters stringens. Det er vores håb, at man i fremtiden vil sørge for at webgrænsefladerne, hvori den danske forskning kan fremsøges, stiller bedre syntakser og kommandosøgningsmuligheder til rådighed.

3. Vi mener at man skal være ganske varsom med at designe informationssøgningssystemers webgrænseflader ud fra et princip om, at de mindst kompetente skal kunne følge med. Som vist mener vi, der er en fare for, at man arbejder sig hen mod laveste fællesnævner. Brugervenlighed og godt interaktionsdesign er selvsagt alfa og omega for en god brugeroplevelse, men disse bør ikke forveksles med lav funktionalitet. Vi opfordrer derfor til, at man udvikler databaser med differentierede webgrænseflader, så de kan betjenes af professionelle såvel som ikke-professionelle informationsbrugere. 🍌