

Informationssøgning set i det Integrerede Kognitive Forskningsperspektiv¹

Af Peter Ingwersen og Kalervo Järvelin

Abstract

Artiklen demonstrerer Laboratorie-perspektivet på informationssøgning og hvordan det er indeholdt i det Integrerede Kognitive Forskningsperspektiv. Først diskuteres Laboratorie-perspektivets underliggende antagelser og velkendte ulemper og begrænsninger. Dernæst diskuteres informationsinteraktion fra et Integreret Kognitivt Perspektiv. 'Ultra-light' og interaktive 'light' eksperimenter med informationssøgning foreslås udført baseret på dette forskningsperspektiv gennem udnyttelsen af perspektivets kontekstuelle muligheder. Disse inkluderer uafhængige variable i relation til samlingen af dokumenter, genfindingsprincipper i søgesystemet og informationssøgerens situation og interessekontekst. Artiklen afsluttes med et konkluderende sammendrag.

Peter Ingwersen er professor ved Danmarks Biblioteksskole. pi@db.dk
Kalervo Järvelin er professor ved Department of Information Studies, University of Tampere, Finland. kalervo.jarvelin@uta.fi

Indledning

Ifølge Järvelin (2007) er begrebet *information retrieval theory* (herfra benævnt informationssøgnings-teori, som også dækker området *information seeking*) sidst nævnt som et nøgleområde i 2000 i de udsendte konferencematerialer fra den mest centrale konference på området: ACM SIGIR. Det indgår i konferencesektionen for 2000 under "IR Theory, including logical, statistical and interactive IR models, and data fusion". Siden da hedder den tilsvarende sektion i invitationen til at indsende bidrag "Formal Models, Language Models, Fusion/Combination". Dermed forstår man hvordan begrebet *teori* fortolkes i en ACM-SIGIR kontekst. Studier af interaktion eller brugere anbringes som regel under en anden sektion, typisk med underkategorier som "Interactive IR, user interfaces, visualization, user studies, and user models" (s. 970).

I denne artikel diskuterer vi udvalgte elementer af en informationssøgningsteori, dvs. Det Integrerede Kognitive Forskningsperspektiv på informationssøgning, som foreslået og analyseret af Ingwersen og Järvelin (2005) idet Laboratorieperspektivet på forskningen også inddrages. De udvalgte elementer er *interaktion* og *relevans*, da disse fænomener er centrale for forskningen og forståelsen af informationssøgning.

Ved *teori* forstår vi systematisk indsamling af teoretiske og empiriske lovmæssigheder og de hermed associerede antagelser og hypoteser. En teori forklarer observerede mønstre og producerer hypoteser om

nye mønstre og årsagssammenhænge. Desuden giver en teori en dyberegående forståelse for fænomener gennem anvendelsen af et begrebsapparat, som går ud over det umiddelbart observerbare. Dette indebærer, at videnskabelige teorier repræsenterer virkeligheden, systematiserer vor viden om den og guider forskningen videre, fx ved at foreslå nye hypoteser (Bunge, 1967; Järvelin, 2007).

Den *integrerede* kognitive teori for informationssøgning opstod 1990-1992 og blev dybere analyseret i (Ingwersen, 1992; 1996). Dette ledte til en stigning i antallet af empiriske studier, baseret på hypoteser, der byggede på denne teoriudvikling (Ingwersen & Järvelin, 2005). Det integrerede perspektiv erstattede en mere individualistisk orienteret tilgang til det 'kognitive synspunkt' på informationsinteraktion, som havde domineret denne forskningsgren siden 1970-erne. Blandt de *moderne* hypoteser finder vi princippet for polyrepræsentation (Ingwersen et al., 2008; Kelly et al., 2005; Larsen, et al., 2006; Skov et al., 2008; White, 2006); alternative forståelser af begreberne 'information' og 'relevans' (Borlund, 2003); og et nyt forskningsperspektiv (Ingwersen and Järvelin, 2005, 313-358). I denne artikel fokuserer vi på de elementer af det integrerede kognitive forskningsperspektiv eller -ramme, som inkluderer den snævre Laboratoriebaserede forskningstradition og de udenom liggende konteksttyper. Det Integrerede Kognitive Perspektiv på forskningen i informationssøgning søger at fremkomme med bredere, men også mere dybtgående hypoteser og forskningsdesign ved at tilføje *kontekst*. Det er målet, at det integrerede forskningsperspektiv leder til et egentligt forskningsprogram for informationssøgning og -formidling (s. 359-376).

Vi er på linie med D.C. Engelbart (1962), når vi anser Laboratiemodellen, Figur 1 og 3, og den kognitive model for informationssøgning, Figur 4-5, som '*conceptual frameworks*' (perspektiver), som specificerer (s. 2):

- „ [Essential] objects or components of the system to be studied.
- The relationships of the objects that are recognized.
- The changes in the objects or their relationships that affect the functioning of the system.
- Promising or fruitful goals and methods of research.”

Vi definerer begrebet *model* på linie med Järvelin (2007) som en præcis (ofte formel) repræsentation af objekter og relationer (eller processer) indenfor et forskningsperspektiv (et 'framework'), eksemplificeret ved begreberne 'sandsynlighedsmodel' og 'vektorrumsmodel' for informationssøgning. I princippet kan en *model* også betegne en modellering af aktører og organisationer, fx som vist i relevansbedømmelses-modellerne, Figur 2 og 6.

Allerede i 1992 startede Robertson & Hancock-Beaulieu diskussionen om de såkaldte „tre revolutioner” indenfor forskningen i informationssøgning: den interaktive, den relevans-relaterede og den kognitive revolution. Den sidste revolution så de som en konsekvens af de to første. Informationssøgning kunne ses som en proces, som leder eller hjælper til menneskelig erkendelse, dvs. opfattelse og fortolkning af information, læring, erkendelse og skabelse af viden. Dette syn på forskningen i (interaktiv) informationssøgning og -adfærd falder sammen med den tids opfattelse af, at denne forskning er opdelt i en 'hovedstrømning', bestående af en traditionel, systemorienteret og laboratoriebaseret linie, også kaldet det algoritmiske perspektiv på informationssøgning eller Cranfieldparadigmet, og det naturalistisk-realistiske, bruger-centrerede, interaktive, synspunkt på informationssøgning (Ingwersen, 1992). Selvom denne binære opdeling af forskningen var ret så klar, bidrog den ikke direkte til at forbedre forskningsmiljøet.

Denne problemstilling kunne ikke løses umiddelbart, hovedsagelig fordi de to poler ikke modtog nogen befrugtning fra hinanden og det syntes umuligt at finde en fælles basis imellem forskningslinierne. Det Integrerede Kognitive Forskningsperspektiv er derfor et forsøg på at bygge en bro imellem de to synspunkter på forskningen til gensidig berigelse (Ingwersen, 1996; Ingwersen & Järvelin, 2005). Årsagen til at dette perspektiv måske kan få succes, og sandsynligvis allerede har bidraget positivt til en mere afslappet holdning i forskningsmiljøet til nye former for forskningsdesign og evalueringsmetoder er, at alle parter er enige om, at det ultimative *mål for informationssøgning* er at lette adgangen til og interaktion med information, og dermed til erkendelse og viden. Samtidig kan man også hævde, at jo bedre systemerne er til at samarbejde med søgeren (og omvendt) for at finde den rette information, jo bedre bliver den overordnede søgesucces for alle parter.

Disse ideer om gensidig udveksling af resultater på tværs af forskningsperspektiver udgør grundlaget for et stigende antal laboratoriestudier og felteksperimenter med såkaldt implicit og eksplicit relevansfeedback; søgeres beskrivelse af deres problem-situation og opgavekontekst; brug af simulering; anbefalingssystemer, social indeksering og søgeadfærd; samt personalisering af søgeprocessen.

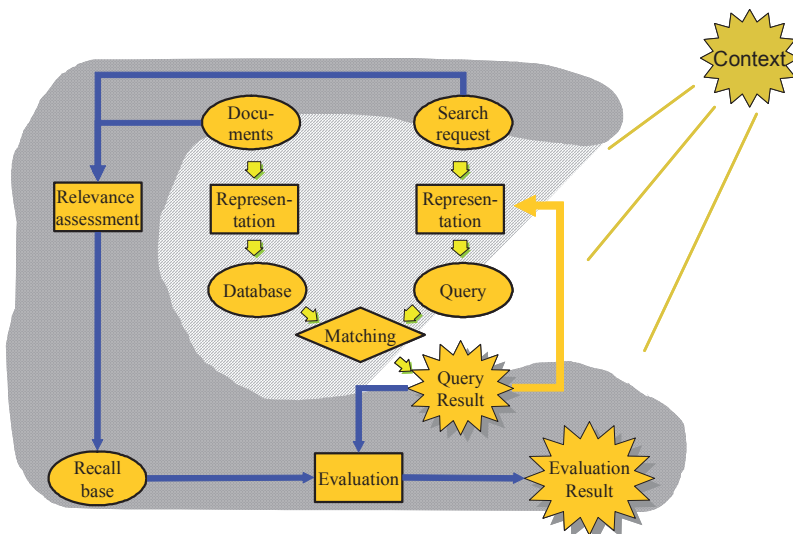
I deres forskningsmonografi opregner og diskuterer Ingwersen and Järvelin (2005) forskellene og sammenfaldene mellem Laboratorieperspektivet og det Integrerede Kognitive Forskningsperspektiv (s. 192-194). Forskellene rækker fra opfattelsen af begrebet information; problem- og opgaveløsnings (u)afhængighed af hinanden; formidlerens rolle; over kontekstens betydning; til evalueringsprocedurer. De to mest centrale forskelle vedrører begreberne *interaktion* og *relevans*. Dette berører forskningsdesign og systemevaluering. Denne artikel forsøger at demonstrere, at de to forskningsperspektiver faktisk hænger tæt sammen og tilsammen bidrager til en bedre forståelse af begge begreber og hele informationsøgeprocessen.

I sektion 2 diskuteres Laboratorieperspektivet, med fokus på dets underliggende antagelser og kendte begrænsninger, men også muligheder. Dette følges af en diskussion af interaktion og relevansfænomener i relation til evaluering af søgesystemer, der har betydning for at se og forstå informationsøg-

ning i en forlængelse af Laboratorieperspektivet hen imod dets kontekst. Sektion 3 diskuterer interaktiv informationsøgning, set fra et Integreret Kognitiv Forskningsperspektiv. 'Ultra-light' og 'light' interaktive søgeeksperimenter foreslås i sektion 4 udført indledningsvis, for at nå ud og ind i de kontekstuelle områder af informationsinteraktionen mellem søgere og system. Disse eksperimentelle situationer illustrerer uafhængige variable fra en dokumentssamling, genfindingsalgoritmer og brugerens situation. Artiklen afsluttes i sektion 5 med et konkluderende sammendrag.

Laboratorieperspektivet på informationsøgning

Laboratorieperspektivet er vist på Figur. 1 i den såkaldte 'grotte-metafor', grundet dets næsten kontekstfrie natur. Ifølge Ingwersen og Järvelin (2005, s. 4-6) består det af et system til informationsøgning, bestående af en dokumentbase, søge- og indekseringsalgoritmer, samt søgeforespørgsler (requests) og systemforespørgsler (queries), og et sæt relevansbedømmelser lavet af et antal bedømmere og lagret i en database for sig (recall base). Systemkomponenterne er placeret i midten af figuren, mens evalueringselementerne ses til venstre og nederst i det grå område. De centrale fokuspunkter for forskningen har enten været repræsentationen af dokumenter og forespørgsler (indekseringsalgoritmer og vægtning) eller søgealgoritmer, som kan producere et relevanssorteret output (query result).



Figur 1. Laboratorieperspektivet på forskningen i informationsøgning. (Revision af Ingwersen & Järvelin 2005, s. 5).

I Laboratorieperspektivet er søgere og brugere af information samt de interesser og behov, der ligger til grund for en søgning ikke nødvendige som led i forskningen. Der er ikke brug for dem til tests af søgealgoritmer op mod hinanden, idet algoritmernes egentlige formål er ret afgrænset: fremfindning og sortering af emnemæssigt relevante dokumenter. Testforespørgsler er typisk emnemæssigt veldefinerede og rige på ord. Dette giver søgealgoritmerne optimale arbejdsbetingelser, men gør også, at forsøgene kan forekomme ret urealistiske; tænk blot på websøgninger med gennemsnitligt to ord per forespørgsel (Spink et al., 2001). Heldigvis er de seneste års globale TREC-eksperimenter, (Text REtrieval Conference) blevet udvidet til også at omfatte en webforsøgslinie med realistiske korte søgeforespørgsler, i nogle forsøg også inddragende egentlige søgere.

Alene at evaluere de algoritmiske komponenter retfærdiggør på sin vis dette forskningsperspektivs måde at bedømme søgekvaliteten på. I dette perspektiv er *målet* gennem konkurrence at udvikle forbedrede søgealgoritmer, som kan identificere og sortere emnerrelevante dokumenter, så de kan præsenteres for en bruger med en given forespørgsel. I Laboratorieperspektivet er forskningen dybest set til for at forbedre *søgesystemers* (algoritmiske) design (Ingwersen & Järvelin, 2005, s. 6-7).

I en teoretisk sammenhæng er målet med denne forskning at analysere grundlæggende problemer i informationsøgning. Det er fx brugen af vokabularier og de sproglige problemstillinger som dukker op med anvendelsen af enkelttermer, sammensatte ord, begreber, semantik, samt repræsentation og matchning af dokumenter og systemforespørgsler samt udviklingen af (formelle) teorier til at løse dem.

Laboratorieperspektivets *eksperimentelle design* er karakteriseret ved bestemte typer af tests, testsamlinger med sæt af forespørgsler og dertilhørende relevansbedømmelser af dokumenter per forespørgsel ('topics' i TREC-sammenhæng), samt etablerede metoder i forbindelse med måling af søgekvalitet, fx recall/precision, MAP (mean average precision) og cumulated gain (Järvelin & Kekäläinen, 2002).

Eksperimenterne er normalt udført i batch-mode. Flere efterfølgende søgninger indenfor samme eksperiment kan forekomme, idet forsøgsopstillingen kan geares til, at et bestemt antal af de højst sorterede

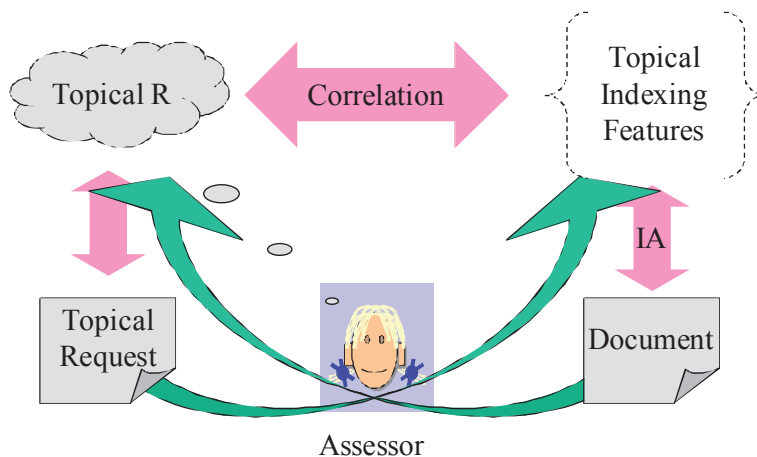
dokumenter automatisk anses for relevante. Disse udgør nu en ny 'forespørgsel' til systemet (tilbagekoblingspilen, Figur 1), som dermed forsøger at forbedre det samlede resultat. Denne type 'pseudo-relevansfeedback' kan efterprøves med forskellige tærskelværdier i forskellige domæne- og mediemæssige sammenhænge og kan ses som en *simulering* af en primitiv informationsøger.

De mest anvendte testsamlinger er for tiden store samlinger af historiske nyhedsdokumenter, fx fra Wall Street Journal og Financial Times. Disse er suppleret med websamlinger og mindre specialsamlinger fra andre medier.

I de seneste år har man afprøvet et slags udvidet Laboratorieperspektiv, Figur 3, ved i TREC at gennemføre et interaktivt spor. Imidlertid afgrænses disse forsøg ofte ved, at de eksisterende relevansbedømmelser indgår, selv om forsøgspersoner foretager søgningerne. Det eksperimentelle forløb kan så blot afprøves henover to sammenhængende søgerunder pr. søgesession. Mere end to runder vil ændre forsøgs karakteren grundet forsøgspersonernes læring under forsøget. Resultatet af en sådan længerevarende søgning vil som konsekvens fjerne sig mere og mere fra den oprindelige relevansbedømmelse udført efter én søgerunde. Denne uhyre korte type interaktive forsøg benævner vi *IR² interaktion 'ultra-light'*.

Betrager vi Figur 1, må vi konstatere at grotteloftets to centrale komponenter, dokumenter og søgeforespørgsler (og deres typer og genrer), kunne undergå flere eksperimenter i dybden end tilfældet hidtil har været, selv *indenfor* Laboratorieperspektivet. Vi har som regel at gøre med 'emne-' eller 'faktuelle forespørgsler', sjældnere med forespørgsler, der inkluderer andre metadata eller blandinger.

Ligeledes må man konstatere, at selve relevansbedømmelsesprocessen ikke udgør et centralt studieobjekt. Undtagelsen er nogle få undersøgelser af, hvordan forsøgsopstillinger gennem statistiske metoder neutraliserer afvigelser fra gennemsnittet, fx hvis man anvender mere end én bedømmer pr. emneforespørgsel (Vorhees, 1998), eller laver sammenligninger mellem binære og skalerede relevansbedømmelser (Sormunen, 2002).



Figur 2. Emnerelevans korrelerer med indekstermer ved emneforespørgsler indenfor Laborierperspektivet (Kekäläinen and Järvelin, 2002a).

Laborierperspektivets Fordele, Ulemper, Antagelser og Mål

En af de centrale fordele ved Laborierperspektivet og den iboende Cranfield Model kan forklares via Figur 2. Enkeltstående nøgleord, og andre eksplicite kendetegn fra tekster eller andre medieformer, korrelerer med høj sandsynlighed med dokumenters emneindhold, når disse kendetegn destilleres af en indekseringsalgoritme (IA) og i stedet resulterer i videnrepræsentationer, samt med forespørgsleres indhold, når disse matches. Undtagelser er problemer med homonymer.

Når en forsøgsperson (eller en algoritme) behandler en emneforespørgsel, er det med stor sandsynlighed muligt at forudsige, hvilke nøgleord (eller kendetegn) der bør trækkes frem (undtagen vedr. problemer med synonymer og parafraser). Fordi emneforespørgsler netop tilgodeser kriterier for emnerelevans, er chancen for en fair korrelation til' stede mellem kendetegnene i de fundne dokumenter og positive relevansbedømmelser (større end ved tilfældige udtræk). Jo flere nøgleord og -kendetegn der kan anvendes som repræsentationer af dokumenter og forespørgsler, des bedre søgekvalitet.

Disse observationer er centrale for forståelsen af Laborierperspektivets succes i tekstbaseret informationssøgning. Alle de udtrukne nøgleord ses som *uafhængige* af hinanden, uanset hvilken søgealgoritme, der anvendes. I de forsøg man har udført med afhængige nøgleord, er der ikke indtruffet mærkbare for-

bedringer af søgeresultatet. Dette viser, at nøglen til søgesucces ligger et andet sted, selv om situationen intuitivt må opfattes som urealistisk: Forfattere skriver jo intentionelt og meningsfuldt. Men det er netop denne meningsfuldhed i teksterne, som leder til succes for informationssøgning: Jo flere meningsfyldte enkelttermer, der bringes sammen i en søgning, jo større er chancen for at ramme nogle få meningsfyldte dele (sætninger) i dokumenter. Få nøgleord rammer derimod uhensigtsmæssigt mange (dele af) dokumenters fulde tekster. Man skal her notere sig, at skriveprocessen ikke er kaotisk men stokastisk, så at sige 'tilfældig', hvis man betragter fordelingen af ord over store tekstmængder (Egghe & Rousseau, 1990). Denne 'tilfældighedsfordeling' gælder imidlertid alene de funktionelle ord i teksterne (Bookstein & Swanson, 1974), *ikke* de indholdsbærende ord. Disse er informative ord, som er bedst anvendelige som indekseringstermer, idet de diskriminerer dokumenterne fra hinanden. Disse ords fordeling er ikke tilfældig, men følger et mønster organiseret af det tematiske forløb i teksten (Katz, 1996, s.16). Der eksisterer således en kognitiv forklaring på, hvorfor de ellers indholdsmæssigt isolerede nøgleord alligevel fungerer så godt i Laborierperspektivets søgeprocesser.

Udvidelsen af Laborierperspektivet til Interaktion 'Ultra-Light'

Figur 3 fremstiller Laborierperspektivet udvidet med forskellige typer implicit eller eksplicit rele-

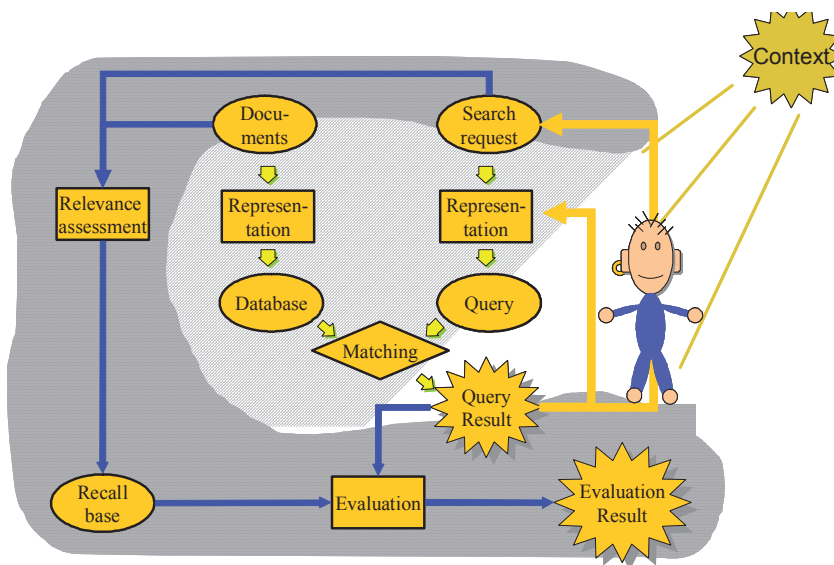
vansfeedback, der er frembragt af en person, som agerer informations søger. Denne lettere udvidelse af perspektivet skyldes ikke mindst opfindelsen af sandsynlighedsmodellen for informations søgning (Robertson, 1977). Denne kræver netop en menneskelig relevansfeedback (højre lodrette pil), for at relevansodds kan beregnes af søgesystemet. Som omtalt ovenfor, tillader Laboratorieperspektivet maksimalt kun to søgerunder pr. søgesession, hvoraf den sidste runde er et resultat af menneskelig relevansfeedback. Hermed bevares de eksisterende relevansbedømmers værdi, og dermed 'Recall-basen' intakt, idet der ikke forekommer nogen læringsprocesser hos aktøren.

Denne udvidelse af Laboratorieperspektivet på IR indebærer, at søgekvaliteten måles efter andet gennemløb op mod det første, automatiske gennemløb, som ses som benchmark. Dette tillader mange eksperimentelle variationer med typer af automatisk og/eller menneskelig relevansfeedback, forsøg med vægtning af nøgleord og kendetegn, og med algoritmisk modifikation af forespørgsler. Der findes nu to måder hvorpå man kan bedømme relevansen af en sådan kort søgesession bestående af to gennemløb. (1) relevans måles som hidtil i de eksisterende testsamlinger én gang for alle af en ekspertbedømmer (ved måling af sæt af dokumenter, fremfundet fra flere søgemaskiner i første gennemløb over samme forespørgsel lavet af bedømmer, som i TREC). (2) forsøgsperso-

ner bedømmer (første og) andet gennemløbs fremfundne dokumenter, på tværs af søgemaskiner og givne forespørgsler. Dette kræver et forskningsdesign, som undgår eller neutraliserer læringseffekter o.l. Første gennemløb findes lagret i en log.

Denne IR interaktion 'ultra-light' situation er på den ene side noget urealistisk (kun to søgerunder), men samtidig optræder relevansbedømmelse (2) som brugerbaseret. Der vil naturligt fremkomme variationer i disse bedømmelser, også i forhold til de oprindelige bedømmelser af eksperterne, som følge af forskellige fortolkninger hos forsøgspersonerne. Netop derfor skal det eksperimentelle design være stringent og robust. I øvrigt vil sådanne 'ultra-light' eksperimenter give anledning til at sammenligne de oprindelige bedømmelser med forsøgspersonernes, samt til at indføre skaleret, ikke-binære relevansbedømmelser (Borlund & Ingwersen, 1998, Borlund, 2000; 2003b; Sormunen, 2002). Den seneste udvikling på området har gjort det muligt at generalisere kvalitetsmålene (Järvelin & Kekäläinen, 2000, 2002; Kekäläinen & Järvelin, 2002b;). Samtidig har TREC-HARD introduceret søgning af højrelevante dokumenter (Vorhees, 2001; Kekäläinen, 2005; Kelly et al., 2005).

Målemetode (1) giver mere konsistente mål, da originalbedømmerne som regel også har genereret emneforespørgslerne (TREC topics). Ved målemetode (2)



Figur 3. Det udvidede Laboratorieperspektiv på informationssøgning ('Ultra-Light').

er dette ikke tilfældet, idet de samme forespørgsler anvendes, men gives til forsøgspersonerne, som nu bedømmer dem. Dette giver en større subjektivitet og dermed usikkerhed, som søges neutraliseret gennem anvendelsen af statistisk tilstrækkelig mange forsøgspersoner og korte søgesessioner. I interaktive 'Ultra-light' laboratorieeksperimenter søger man at bevare kontrollen over eksperimenterne, men fjerner sig fra realismen i forsøget.

Man kan undgå fortolkningsproblemet og usikkerheden ved at *simulere* informationssøgere i disse 'ultra-light' forsøg. Dette indebærer, at man enten via informationssøgelitteraturen eller på anden måde danner hypoteser om personers søgeadfærd, fx i forbindelse med relevansfeedback. I forsøg opstilles alle de mulige kombinationer af adfærdsmønstre, som hypotesen leder frem til. Disse afprøves, evt. over mange gennemløb. Da ingen forsøgspersoner deltager, indtræffer der ingen fordrejninger i resultatfasen. Den eller de bedste kombinationer af adfærd, set fra et søgekvalitetssynspunkt, kan senere afprøves i et 'ultra-light' eksperiment med forsøgspersoner eller i interaktion 'light' eller naturalistiske forsøg. Der er lavet en del simuleringseksperimenter. Magennis og Van Rijsbergen (1997) gjorde forsøg med automatisk modifikation af forespørgsler, White et al. (2004; 2005) eksperimenterede med implicit relevansfeedback og brugervalidering, Wang et al. (2006) kiggede på logbaseret 'collaborative filtering' og Keskustalo et al. (2006) undersøgte skaleret relevansfeedback.

Vi er nu på vej ind i det Integrerede Kognitive Forskningsperspektiv på informationssøgning, hvor forsøgspersonerne modtager en simuleret informations-situation (en såkaldt dækhistorie), som opstart til et relativt kontrolleret forsøg, eller de producerer selv et informationsbehov (Borlund, 2003b). I det sidste tilfælde er forsøget realistisk, men ret ukontrolleret. Gennem anvendelsen af en simuleret søgesituation opnår man en balance mellem realisme og kontrol, idet situationen er åben for fortolkning (realisme), men samme situation gives til mange forsøgspersoner (kontrol). I dette *IR interaktion 'light'* forsøg vil målemetode (2) altid blive brugt, idet en testsamlings oprindelige bedømmelser altid forudsætter ikke-dynamiske informationsbehov og statisk relevans. Samtidig forudsætter Laboratorieperspektivet, i modsætning til det Integrerede Kognitive Forskningsperspektiv, at de oprindelige bedømmere aldrig lider af

'træthed' og altid 'glemmer' de foregående bedømmelser.

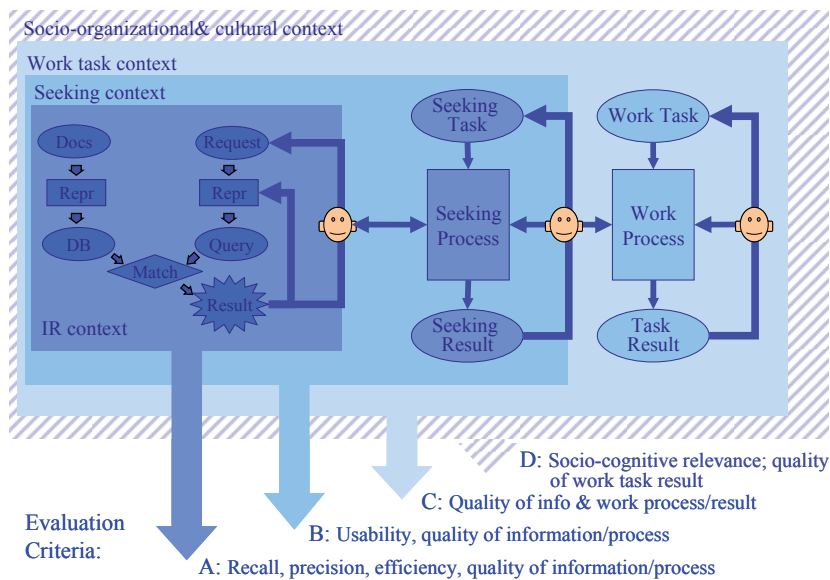
I de såkaldte INEX forsøgsrækker (Initiative for the Evaluation of XML retrieval) har man, i modsætning til TREC, anvendt simulerede situationer som baggrund for eksperternes skalerede relevansbedømmelser. INIX er dermed bedre gearret til interaktionseksperimenter med forsøgspersoner (Malik et al., 2007). Antallet af søgerunder vil være bestemmende for, om et forsøg tilhører interaktion 'ultra light' eller 'light'.

I Laboratorieperspektivet opereres der med de samme uafhængigheds-antagelser i forbindelse med selve *interaktionen* som gælder for kendetegn i dokumenter og relevansbedømmelser. Den realistiske, komplekse, og dynamiske interaktion, som vi kender fra dagliglivet, er her brudt op i enkeltstående og uafhængige interaktionsmomenter. En succesfuld enkeltstående interaktion bliver derfor belønnet i dette forskningsperspektiv. Skulle en aktør skifte fortolkning af sin situation, fx på grund af de observerede dokumenter, opstår der blot en *ny* søgesituation og en *ny* enkeltstående interaktion. Denne måde at anskue interaktionsforløb på er ikke blot urealistisk, men på sin vis også i strid med ideen bag relevansfeedback i forbindelse med sandsynlighedssøgning, nævnt overfor, som netop indebærer et sammenhæng mellem søgerunderne i interaktionen. Dette punkt er sjældent behandlet indenfor Laboratorieperspektivet.

Det Integrerede Kognitive Perspektiv på IR

Som omtalt ovenfor nærmer Laboratorieperspektivet på informationssøgning sig nødvendigvis den kognitive og fysiske kontekst udenfor laboratoriegrottens metaforiske åbning, Figur 4.

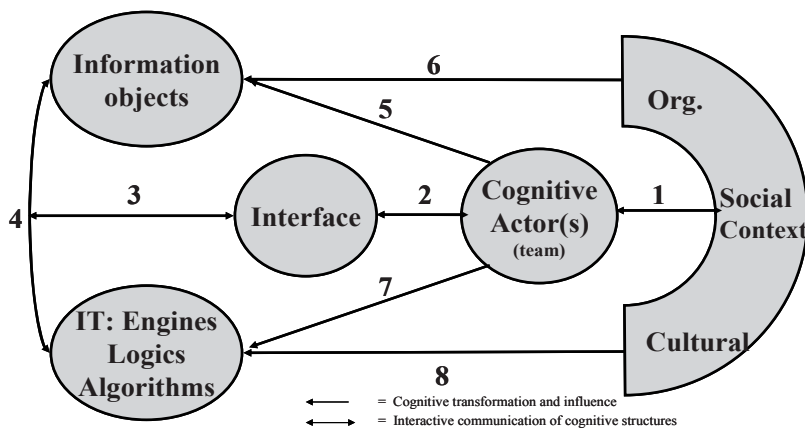
Algoritmisk informationssøgning ses her i kontekst af bredere informationssøgeadfærd samt opgavesituationer og –processer, jobbaserede eller fra dagliglivet (Ingwersen & Järvelin, 2005, s. 322). Interaktive processer finder sted horisontalt, mens relevansbedømmelser og –feedback sker vertikalt. For hver ny konteksttype angiver modellen typerne af bedømmelses-kriterier, som kan anvendes i forsøgsoptillinger. Modellen i Figur 5 er en videreudvikling af det Integrerede Kognitive Forskningsperspektiv. De centrale antagelser er følgende (s. 25):



Figur 4. Konteksttyper og bedømmelseskriterier for opgavebaseret informationsadgang (udvidelse af Kekäläinen & Järvelin, 2002a)

1. Informationsbehandling finder sted hos både afsendere og modtagere af meddelelser;
2. Informationsbehandlingen sker på forskellige sproglige niveauer;
3. Under kommunikation af information er enhver aktør påvirket af sin fortid og nuværende erfaringer (tiden) og den organisatorisk-sociale og kulturelle kontekst;
4. Individuelle aktører påvirker omgivelserne og domænet;
5. Information er situationel and kontekstuel.

For det Integreerede Kognitive Perspektiv på forskningen i informationsøgning betyder det, at den informationsbehandlede mekanisme kan optræde som afsender eller som modtager af tegn, signaler eller data, gennem en kommunikationsproces. Dette indebærer, at perspektivet ikke blot ser menneskelige aktører som modtagere, men også som ophav til tegn til og fra datamaskiner og vidensobjekter (pil 5-8) i Figur 5. Der er således altid flere forskellige menneskelige aktører på spil i IR som variable – ikke blot designere af algoritmer, relevansbedømmere i testsamlinger og en informations søger.



Figur 5. Det Integreerede Kognitive Perspektiv på informationsøgning (Ingwersen & Järvelin, 2005, s. 261).

Det Integrerede Kognitive Forskningsperspektiv er derfor ikke afgrænset til de bruger-orienterede tilgange til information og informationssøgning. Det favner alle informationsbehandlende aktører og mekanismer skabt af mennesker for mennesker. Data-maskiner og anden form for teknologi indgår derfor på linie med ophav, brugere og diverse processer som overføring og perception samt fortolkning af information.

Venstre side af modellen inddrager Laboratorieperspektivet på informationssøgning, jf. Figur 4 og illustrerer med pil 2 søgeforespørgslen og relevansfeedback. Pil 3 henviser til systemforespørgslen (query) eller dens modifikation, mens pil 4 illustrerer indhentningen af dokumenter ved hjælp af en søgealgoritme. Pil 1 refererer til den sociale interaktion mellem fx en informationssøger og dennes umiddelbare sociale kontekst, dvs. informationssøgning uden inddragelse af formelle søgesystemer, dvs. uden IR. Alle fem komponenter i modellen, Figur 5, udgør kontekster til hinanden. Desuden optræder *interaktion* som en sjette central komponent i det Integrerede Kognitive Forskningsperspektiv. Man kan ikke betragte informationssøgeren isoleret fra sin sociale/organisatoriske kontekst; ej heller kan man se aktøren uden en *systemkontekst*, Laboratorieperspektivet så at sige. Når én komponent skifter indhold eller form, påvirker det de andre komponenter og kontekster. Der er konstant en dynamisk ubalance til stede i interaktionsprocessen.

Interaktion og relevans

Som følge af den opsplittede søgeinteraktionsproces under Laboratorieperspektivet er der stort set ingen forsøg med søgestrategier inden for dette forskningsperspektiv. De forskellige blokbygningsteknikker kendt fra traditionel online søgning ved hjælp af sætkombinationer kan ikke overføres til de algoritmiske søgemodeller. Dette gælder også i forbindelse med Google og Yahoo, som teknisk set er Boolsk søgning i ultra-terabyte skala med en relevanssorteringsalgoritme lagt ovenpå. De samme urealistiske uafhængighedsantagelser, som vi omtalte ovenfor, gælder imidlertid også her: en forandring i søgeprofilen indebærer en ny isoleret søgning, og den foregående er glemt af maskinen.

Det er først i de senere år lykkedes at (gen)indføre strukturerede søgeforespørgsler, som dermed udnytter blokbygnings- og facetprincipper i den algoritmi-

ske IR (Kekäläinen & Järvelin, 1998). Det har vist sig, at denne form for struktur forbedrer søgeresultatet i forhold til brugen af isolerede nøgleord i de avancerede søgealgoritmer.

I naturlig interaktion (og ved brug af simulerede informationssituationer) ses det ofte, at forsøgspersoner fremkommer med forespørgsler, hvis søgeresultat er meget dårligt. Realistiske forespørgsler er som regel korte. Relevansfeedback hjælper ikke altid, da personens viden på området kan være svag. Dette samt det faktum, at flere forsøgspersoner som regel fortolker de samme situationer forskelligt, gør at IR interaktion 'light' og forsøg med naturalistisk informationssøgning nødvendigvis må involvere flere end én variabel, som regel mindst to. Mange andre variable skal endvidere kontrolleres eller neutraliseres, for at opnå valide resultater.

I forbindelse med opfattelsen af *relevans* i det Integrerede Kognitive Forskningsperspektiv Figur 4-5, har Borlund (2003a), Cosijn & Ingwersen (2000) samt Saracevic (1996; 2007a; 2007b), foreslået og diskuteret en relevanstypologi. Denne består af relevans af lavere orden, algoritmisk og emnerrelevans, og relevans af højere orden: 'pertinence', situationel relevans og socio-kognitiv relevans. 'Pertinence' betegner relationen mellem informationsbehovet (som fortolket af den enkelte person) og de fundne dokumenter. Kriterier er her typisk relateret til metadata, som publikationstidspunkt, nyhedsværdi for personen, dokumentets ophavsautoritet, o.l. Hvor emnerrelevans (topicality) har at gøre med dokumenters *aboutness*, har 'pertinence' at gøre med deres *isness* (Ingwersen & Järvelin, 2005, s. 271), altså en 'værenhed'. Socio-kognitiv relevans illustreres bedst ved objektive bedømmelser foretaget over tid, fx modtagne citationer eller indlænker, som kan betegne anerkendelse, eller andre former for fagfællebedømmelser. Situationel relevans betegner relationen mellem personens problemsituation (eller interesse og opgaveløsning) og de fundne dokumenter.

Begrebet situationel relevans er vanskeligt at operationalisere. Et eksempel på denne relevanstype er, når en forfatter giver et andet værk en reference (som senere bliver til en modtaget citation) eller skaber en udlænke. Her er det situationen i det pågældende værk, som dikterer relevansen af det refererede værk. Man kan dermed antage, at en artikels referenceliste betegner en liste af situationelle relevansbedømmel-

ser, mens ankerteksten inde i artiklen angiver hvorfor disse referencer er lavet. Denne form for situationel relevans er blevet udnyttet af Schneider (2005; 2006) til at konstruere synonymtesauri og navngive klynger.

Men ellers har denne type relevans det med at være usynlig i undersøgelser af relevanskriterier (Berry and Schamber, 1998; Vakkari, 2001). Emnerelevans er den mest forekomne type, men afhænger af medie, genre og domæne. For web-informationssøgning har Tombros et al. (2003; 2005) fundet, at indholds- og layoutkriterier udgør de mest benyttede bedømmelseskriterier. De fandt ingen situations-beskrivelser, undtagen udtryk som: "this is hot, man". Det der ser ud til at indtræffe for den situationelle relevans er, at personen netop anvender nogen af de samme kendetegn ved dokumentet, som også normalt anvendes ved emnerelevans (Burgin, 1992; Vakkari, 2001). Dermed *kollapser* den situationelle relevans under forsøgene, og ender som lavere ordens emne- eller indholdsrelevans. Noget tyder på, at ved at forbedre indekseringen af dokumentet gennem human indeksering rettet direkte mod dokumentets brug i en arbejdsopgave-sammenhæng, kan man fremtvinge en situationel relevansbedømmelse hos en aktør under søgeprocessen.

Dette fænomen ligner det såkaldte '*kognitive frie fald*', som diskuteres i kognitiv IR teori: at kommu-

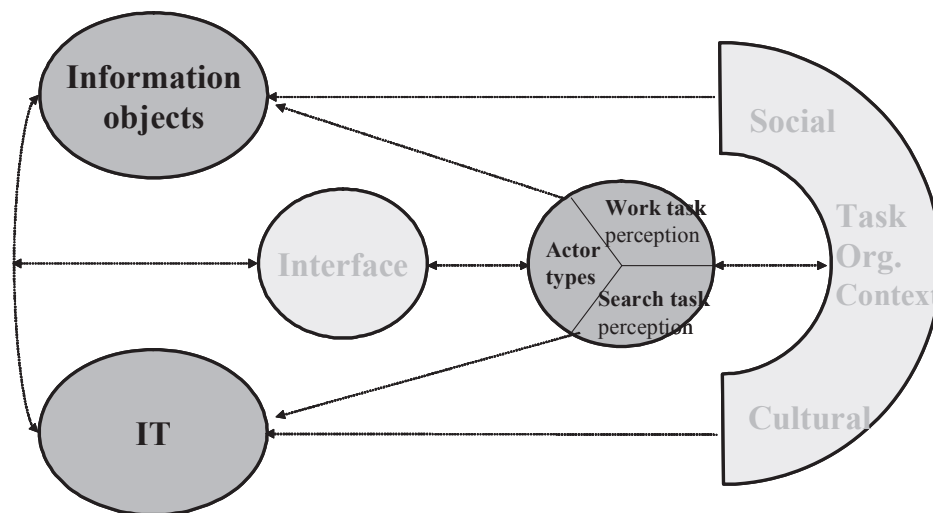
nikerede meddelelser taber al mening og intention og disintegrerer i enkeltstående tegn (Ingwersen & Järvelin, 2005, s. 33-38)

Udenfor Laboratorie-grotten og ind i kontekst: Et Illustrativt Eksempel

Ved at bevæge sig ud i de kontekstuelle komponenter i det Integrerede Kognitive Forskningsperspektiv er man tvunget til at lave forsøgsopstillinger, som er så kontrollerede som muligt, samtidig med at realismen bevares (Ingwersen & Järvelin, 2005, s. 313-376). Dette kan illustreres ved et udvalgt eksperimentelt eksempel nedenfor. Eksemplet illustrerer hvad man kan kalde 'naturalistisk' IR interaktion. Udføres forsøgene i et laboratorium med simulerede søgeopgaver under meget kontrollerede forhold og med 1-2 uafhængige variable, benævnes eksperimenterne IR interaktion 'light'.

I alt indeholder forskningsperspektivet 9 dimensioner af variable, hver med deres værdier. Dimensionerne hidrører fra de seks komponenter i Figur 5, inkl. interaktionsprocessen. Aktøren(e) består af tre dimensioner:

- personlige egenskaber;
- fortolkning af herværende situation mhp. arbejdsopgaveløsning eller interesse;
- fortolkning af søgeopgaven.



Figur 6. Dimensioner af det Integrerede Kognitive Forskningsperspektiv, hvorfra uafhængige variable indgår i en forsøgsopstilling i interaktiv informationssøgning (mørke cikler). (Ingwersen & Järvelin, 2005, s. 364)

Desuden består den social-organisatoriske og kulturelle kontekst af to dimensioner af variable:

- naturlige arbejdsopgaver eller interesser;
- naturlige søgeopgaver.

Figur 6 viser et forsøgsscenarie, hvor de uafhængige variable vedrører informationsaktør, informationsobjekter og IT-komponenten i en informationssøgesituation.

Forskningsspørgsmål og Typen af Variable i eksemplet

Denne forsøgsopstilling indeholder Laboratorieperspektivet, men udvidet til at omfatte informationsobjekten i en naturlig, længerevarende interaktion med søgesystemet, Figur 6. Forskningsspørgsmålet kunne være:

Givet en specifik organisation X med kendte arbejdsopgaver, hvilken søgealgoritme arbejder så bedst i forhold til to forskellige videnressourcer, når vi tager i betragtning, at søgerne har forskellige erfaringsgrundlag ved udførelsen af arbejdsopgaver?

Organisationen kan være et hospital og det faglige domæne klinisk medicin. Denne kontekst bliver dermed en kontrolleret variabel. Typiske arbejdsopgaver er at stille diagnoser, udføre behandlinger, lave kliniske tests, udføre kirurgiske indgreb, udskrive recepter, o.l. Der er tre uafhængige variable i scenariet.

1. *Søgealgoritmerne*, man ønsker afprøvet i dette feltforsøg kunne være vektorrummodellen op imod en netværksbaseret metode.
2. *Informationssøgerne* kan være erfarne overlæger eller 1. reservelæger sat op mod 1. års medicinstuderende.
3. *Videnressourcerne* kan bestå af medicinske fuldttekstartikler versus akademiske websider.

Søgeopgaverne stillet til forsøgspersonerne kunne dels bestå af rigtige arbejdsopgaver, som forsøgspersonerne selv valgte. Men dermed mistes kontrollen let over forsøgene. En mere oplagt mulighed er at udforme realistiske, simulerede arbejdsopgaver, som gør at forsøgspersonerne selv udformer informationsbehov (Borlund, 2003b).

Document and Source types	Algorithmic IT Component	Actor Characteristics	Perceived Work Task	Perceived Search Task
Doc. Structure	Doc. Metadata rep.	Domain Knowledge	Structure/Openness	Inform. Need Types
Doc. Types	Doc. Content rep.	IS&R Knowledge	Strategy/Practice	Structure/Type
Doc. Genres	Doc. Structural rep.	Work Task Exp.	Granularity/Size	Strategy/Practice
Information Types	Req. Metadata rep.	Search Task Exp.	Dependencies	Complexity/Specific
Comm. Function	Req. Content rep.	Work Task Stage	Requirements	Dependencies
Sign Language	Rep. Structural rep.	Context Perception	Domains/Context	Stability
Layout & Style	Match Methods	Constraints	...	Domains/Context
Doc. Isness	...	Motivat./Emotion
Lunk Structures
Human Source

Tabel 1. Uafhængige variable (mørk boks, indrammet), og kontrollerede variable (lysegrå) i et informationsøgningsforsøg.

I tilfælde af simulerede arbejdsopgaver, kan man vælge semantisk set lukkede typer, som fortrinsvis leder frem til faktuelle informationsbehov. De kan bestå af røntgen fotos eller videooptagelser optaget med mikrokamera, som det realistisk foregår på et hospital ved diagnosticering. Alle forsøgspersonerne får stillet de samme simulerede opgaver på tværs af de to typer kilder og søgealgoritmer. Deltagerne i forsøget udfører deres søgninger som de normalt ville gøre det på hospitalet uden afsat tidspres, så vidt muligt, men man forbyder kommunikation til kolleger i forsøget for ikke at tilføre yderligere variable (Ingwersen & Järvelin, 2005, s. 365-366).

Motivationen bag forsøget er den antagelse, at de traditionelle akademiske artikler fremstår som bedre informationskilder for de erfarne læger end websider. Desuden ønsker man at finde ud af hvilken søgealgoritme, som finder de mest relevante dokumenter på tværs af forsøgsgruppen.

Vi skal bemærke, at variabelen 'domæneviden' ligeledes optræder som uafhængig variabel på aktørsiden i Tabel 1. Dette skyldes, at der ofte kan være en tæt korrelation mellem domæneviden og aktørerfaring med problemløsning. Men i nogle tilfælde er de uafhængige af hinanden. Erstattes de studerende i forsøget af sygeplejersker kan man forestille sig, at både læger og sygeplejersker har tilsvarende erfaringsgrundlag, men er på forskellige niveauer i relation til domæneviden. I så fald vil der optræde i alt fire uafhængige variable i forsøget, hvilket realistisk set påtvinger en reduktion til tre ved at lade informationsressourcerne udgøre en og samme type.

Eksperimentel Opstilling

Tabel 1 demonstrerer de tre uafhængige variable, udvalgt fra de tre centrale dimensioner i Figur 6 (mørke, indrammede bokse). Hver variabel kan udtrykkes med forskellige værdier. For eksempel kan variablen 'Document Type' for dimensionen 'Document and Source Types' strække sig fra aviser over monografier til tidsskriftsartikler, conferencebidrag, musikoptagelser, websider, etc. I vort tilfælde optræder to værdier, som nævnt ovenfor. I det foreliggende forskningsspørgsmål er de kontrollerede variable 'Structure/Openness' og 'Domain Context' for dimensionen 'Perceived Work Task' hospitalsomgivelserne, som forsøgene afholdes i samt de givne arbejds- og søgeopgaver, som alle findes indenfor det medicinske domæne. Disse bokse er grålige. Variable 'Information Need Types' (faktuelle informationsbehov) og 'Human Sources' (forbud mod kontakt til andre) er ligeledes kontrollerede. Brugergrænsefladen følger de to algoritmer, som afprøves i forsøget.

Den *afhængige variabel* er søgekvalitet, målt som Recall og Precision (Mean Average Precision = MAP). Alle andre variable i de andre dimensioner (ikke alle vist på Tabel 1) optræder formelt som skjulte variable, som kan lave interferens i forsøgene. Gennem at gøre brug af adskillige forsøgspersoner søger man at *neutralisere* eventuelle mindre forskelle mellem disse, fx vedrørende søgeerfaring.

I denne forsøgsopstilling optræder tre variable, hver med to værdier. Dette er en kompleks forsøgs-situation, som kræver at der gennemføres mange søgninger, for at resultaterne er statistisk valide. I den foreliggende forsøgsrække tænkes at indgå 32 forsøgspersoner (16 læger og 16 medicinske studerende), otte simulerede søgeopgaver (Q1-Q8), to algoritmer under test (a og b), og to dokumenttyper (D1 og D2). Alle disse personer, opgaver, algoritmer og dokumenttyper blandes i forsøget på en meget systematisk og symmetrisk facon:

Otte forsøgspersoner (læger) og 8 studerende får hver to søgeopgaver (Q1-2), som afprøves på algoritme (a) og dokumenttype (D1); desuden afprøver de samme forsøgspersoner (Q3-4) på algoritme (a) og dokumenttype (D2). Dernæst afprøver *den samme* forsøgsgruppe (Q5-6) og (Q7-8) på algoritme (b) og henholdsvis (D1) og (D2).

Otte andre læger og studerende udfører nu forsøgsrækken igen, dog således at de to grupper personer først søger (Q5-Q8) på algoritme (a) og henholdsvis (D1/D2) og dernæst afprøver (Q1-Q4) på algoritme (b) og (D1/D2). Som man bemærker, er der fuld symmetri i udførelsen og hver forsøgsperson kommer igennem alle otte søgeopgaver og begge algoritmer og informationskilder, uden at gentagelser forekommer.

Med denne forsøgsopstilling opnår man dels at forsøgspersonerne er i stand til at udføre søgningerne uden at trættes unødigt (fx over to sessioner á fire forsøg), dels at der ved krydstablering mellem variable optræder 32 tilfælde for hver kombination. Dette sikrer sikkerheden i fortolkningen af resultaterne. I alt gennemføres 256 tilfælde (32 personer gange 8 søgeopgaver).

Det er klart, at ved at reducere antallet af uafhængige variable til to, behøver forsøgsopstillingen færre forsøgspersoner (2 x 8), for at opnå et minimum af statistisk sikkerhed ved fortolkningen. Man kunne derfor tænke sig at udføre forsøgene blot med de to søgealgoritmer og to persongrupper.

Ved at udføre forsøgene i et laboratorium opnås bedre observationsforhold og større kontrol. Et sådan eksperiment udenfor hospitalsmiljøet må dog betegnes som et IR interaktion 'light' forsøg.

Denne forsøgsopstilling gør det også muligt at studere søgeadfærden hos de forskellige grupper. Hvis man ikke kan skaffe tilstrækkelig med forsøgspersoner, er det nødvendigt at hver person udfører mere end 8 søgninger. Med færre personer bliver resultaterne af adfærdsstudierne dermed mindre sikre, grundet den manglende spredning af personer.

Konkluderende sammenfatning

Det Integrerede Kognitive Forskningsperspektiv, som danner rammen om dets ni dimensioner af variable, tilbyder en bredere og dybere begrebsmæssig forståelse af informationssoegeprocessen, samt tillader mere interessante hypoteser end Laboratorieperspektivet eller diverse bruger-centrerede tilgange. Ingwersen og Järvelin (2005, s. 359-380) peger på følgende centrale parametre, som indgår i deres forståelse af interaktiv IR og de fænomener, som knytter sig hertil:

- “[Real] users do not necessarily have ready made verbose needs;
- Real users interpret their situations differently – even if put into the “same” situation (as far as that is possible, e.g., by means of simulated task situations);
- Even if experienced professionals, they may approach the information problems from different angles;
- Real users therefore construct quite different queries and are for long known to assess document relevance differently (Cleverdon, 1984);
- Real users face vocabulary problems and do not know the collection well; therefore initial queries may fail badly and (pseudo) RF [relevance feedback] may not work.
- Therefore, they may issue many consecutive queries on the same (but evolving) topic/need as they learn on the fly what works and what does not.
- Consequently, while ranking well for any given query is important, it is at least equally important to help the user to arrive at a good query (and if it is really good, any ranking method works well).”

Det Integreerede Kognitive Forskningsperspektiv anviser, hvordan man kan behandle de ovennævnte fænomener:

- Anvendelsen af forsøgsopstillinger som det foreslåede i sektion 4 eller andre opstillinger, der kombinerer forskellige variable fra de ni dimensioner i Forskningsperspektivet. Det foreslår som maksimum at anvende tre uafhængige variable parvis, hver med to værdier;
- Sammenligning af informationssøgning i forskellige typer af dokumentsamlinger;
- Sammenligning af fagekspert og novicer (og grupperinger derimellem), som vist ovenfor;
- Sammenligning af naturlige arbejdsopgavers løsning gennem informationssøgning med simulerede opgavesituationer. De sidste kan manipuleres bevidst til at være mere eller mindre semantisk åbne, mens de naturlige opgavesituationer kan anvendes som en kontrolmekanisme overfor de simulerede (Borlund, 2003b).

I denne artikel har vi analyseret og diskuteret Laboratorieperspektivet på informationssøgning og vist hvordan det fuldt indgår i det Integreerede Kognitive Forskningsperspektiv. Vi har ikke blot opregnet de centrale problemer og fænomener, som skal under-

søges og løses, men også søgt et pege på løsninger baseret på teoretiske rammer og gennemarbejdede forskningsmodeller. Vi har vist hvordan det integrerede forskningsperspektiv kan lede til forsøgsopstillinger, som er realistiske og kontrollerede på samme tid. Man kan, efter forskningens målsætning, bevæge sig fra IR interaktion ’ultra-light’ over ’light’ til naturalistisk informationsinteraktion.

Det er via eksperimenter i laboratoriet og undersøgelser i felten, at man opnår forståelse af hvordan de bedste søgealgoritmer faktisk virker i det virkelige liv. Samtidig opnår man viden om menneskers søgeadfærd i forskellige situationer og kontekster.

Det er centralt at finde frem til de mest formålstjenlige ’greb’ i de interaktive søgeprocesser og –systemer; samt hvordan mennesker faktisk håndterer dem. Design af brugergrænseflader og evalueringsmetodik bliver derfor af central betydning for udviklingen af forskningen indenfor informationssøgning i fremtiden. Tendensen gå mod en mere personlig søgeproces, baseret på omfattende social viden om denne proces. Implicit og eksplicit relevansfeedback samt automatisk modifikation af søgeforespørgsler under søgningen (som jo ikke findes i Google eller Yahoo) indebærer grænseflader, som definerer en *balance* mellem den enkelte søgers behov og systemets behov for data om personen. At kunne fange en ’god’ førstegangs-søgeforespørgsel, eller hjælpe til at brugerne kan skabe bedre forespørgsler, udgør væsentlige bidrag til fremtidens IR systemer og digitale medier.

Anerkendelse

Forfatterne ønsker at takke Nordic Research School of Library and Information Science (NORSLIS) for støtte til deres forskningsprojekt.

Noter

1. Denne artikel er en forkortet dansk oversættelse af originalbidraget: On the holistic cognitive theory for information retrieval: Drifting outside the cave of the Laboratory Framework. Ingwersen, P. & Järvelin, K. In: Dominich, S & Kiss, F. (eds.): *Studies in Theory of Information Retrieval*. Budapest: Foundation for Information Society, 2007: 135-147.
2. IR er forkortelsen for Information Retrieval

Referencer

- Barry, CL & Schamber, L (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing and Management*, 31(2/3), 219-236.
- Bookstein, A & Swanson, D (1974). Probabilistic models for automatic indexing. *Journal of the American Society for Information Science*, 25(5), 312-318.
- Borlund, P (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71-90
- Borlund, P (2003a). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913-925.
- Borlund, P (2003b). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), paper no. 152 <http://informationr.net/ir/8-3/paper152.html>. Cited May 13, 2003.
- Borlund, P & Ingwersen, P (1998). Measures of relative relevance and ranked half-life: Performance indicators for interactive IR. In: WB Croft, et al. (Eds.), *21st ACM-SIGIR Conference*. ACM Press, 324-331.
- Bunge, M (1967). *Scientific Research*. Springer. Heidelberg.
- Burgin, R (1992). Variations in relevance judgments and the evaluation of retrieval performance. *Information Processing and Management*, 28(5), 619-627.
- Cleverdon, CW (1984). Optimizing convenient online access to bibliographic databases. *Information Services and Use*, 4, 37-47.
- Cosijn, E & Ingwersen, P (2000). Dimensions of relevance. *Information Processing and Management*, 36, 533-550.
- Egghe, L & Rousseau, R (1990). *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Elsevier. Amsterdam.
- Engelbart, D (1962). *Augmenting Human Intellect: A Conceptual Framework*. Stanford Research Institute. Menlo Park, CA.
- Ingwersen, P (1992). *Information Retrieval Interaction*. Taylor Graham. London
- Ingwersen, P (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50.
- Ingwersen, P & Järvelin, K (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer. Heidelberg.
- Järvelin, K (2007). An analysis of two approaches in information retrieval: from frameworks to study designs. *Journal of American Society for Information Science and Technology*, 58(7), 971-986.
- Järvelin, K & Kekäläinen, J (2000). IR evaluation methods for retrieving highly relevant documents. In: *23rd ACM-SIGIR Conference*. ACM Press, 41-48.
- Järvelin, K & Kekäläinen, J (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (ACM TOIS)*, 20(4), 422-446.
- Katz, S (1996). Distribution of content words and phrases in text and language modelling. *Natural Language Engineering*, 2(1), 15-60.
- Kekäläinen, J (2005). Binary and graded relevance in IR evaluations - Comparison of the effects on ranking of IR systems. *Information Processing and Management*, 41(5), 1019-1033.
- Kekäläinen, J & Järvelin, K (1998). The impact of query structure and query expansion on retrieval performance. In: Croft, WB, et al. (Eds.), *21st ACM-SIGIR Conference*. ACM Press, 130-137.
- Kekäläinen, J & Järvelin, K (2002a). Evaluating information retrieval systems under the challenges of interaction and multi-dimensional dynamic relevance. In: Bruce, H et al. (Eds.), *The CoLIS 4 Conference*. University of Washington. 253-270.
- Kekäläinen, J & Järvelin, K (2002b). Using graded relevance assessments in IR evaluation. *Journal of*

- the American Society for Information Science and Technology, 53(13), 1120-1129.
- Kelly, D, Dollu, VD & Xin Fu (2005). The loquacious user: A document-independent source of terms for query expansion. In: *28th ACM SIGIR Conference*, ACM Press, 457-464.
- Keskustalo, H, Järvelin, K & Pirkola, A (2006). The Effects of Relevance Feedback Quality and Quantity in Interactive Relevance Feedback: A Simulation Based on User Modeling. In: Lalmas, M and al. (Eds.), *28th European Conference on Information Retrieval ECIR '06*, London, April 2006. Heidelberg: Springer, 191-204. Lecture Notes in Computer Science vol. 3936.
- Larsen, B, Ingwersen, P & Kekäläinen, J (2006). The polyrepresentation continuum in IR. In: Ruthven, I. et al. (eds.), *Information Interaction in Context: IliX*. Royal School of LIS, Copenhagen, 148-162.
- Magennis, M & van Rijsbergen, CJ (1997). The potential and actual effectiveness of interactive query expansion. In: Belkin, NJ, Narasimhalu, AD, and Willett, P (Eds.), *20th ACM-SIGIR Conference*. ACM Press, 324-332.
- Malik, S, Tombros, A & Larsen, B (2007). The Interactive Track at INEX2006. In: Fuhr, N., Lalmas, M. and Trotman, A. eds. *Comparative Evaluation of XML Information Retrieval Systems, 5th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006*. Berlin: Springer, 387-399. Lecture Notes in Computer Science ; 4518.
- Robertson, SE (1977). The probability ranking principle in IR. *Journal of Documentation*, 33(4), 294-304.
- Robertson, SE & Hancock-Beaulieu, M (1992). On the evaluation of IR systems. *Information Processing and Management*, 28(4), 219-236.
- Saracevic, T (1996). Relevance reconsidered '96. In: Ingwersen, P. and Pors, N.O. (Eds.), *2nd CoLIS Conference*. Royal School of LIS, Copenhagen, 201-218.
- Saracevic, T (2007a). Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of American Society for Information Science and Technology*, 58(13), 1915-1933.
- Saracevic, T (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effect of relevance. *Journal of American Society for Information Science and Technology*, 58 (13), 2126-2144
- Schneider, J & Borlund, P (2005). A bibliometric based semi-automatic approach to identification of candidate thesaurus terms: Parsing and filtering of noun phrases from citation contexts. In: Crestani, F and Ruthven, I (eds.) *Context: Nature, Impact, and Role - Proceedings of the Fifth International Conference on Conceptions of Library and Information Science (CoLIS 5)*. 226-237.
- Schneider, J (2006). Concept symbols revisited: Naming clusters by parsing and filtering of noun phrases from citation contexts of concept symbols. *Scientometrics*, Vol. 68, No. 3, 573-593.
- Skov, M, Larsen, B & Ingwersen, P (2006). Inter and intra-document contexts applied in polyrepresentation. In: Ruthven, I et al. (eds.), *Information Interaction in Context: IliX*. Royal School of LIS, Copenhagen, 163-170.
- Skov, M, Larsen, B and Ingwersen, P (2008). Inter and intra-document contexts applied in polyrepresentation for best match IR. *Information Processing & Management*, 44(5), 1673-1683.
- Sormunen, E (2002). Liberal relevance criteria of TREC – Counting on negligible documents? In: Beaulieu, M et al (Eds.), *25th ACM-SIGIR Conference*, 320-330.
- Spink, A, Wolfram, D, Jansen, BJ & Saracevic, T (2001). Searching the Web: the public and their queries. *Journal of the American Society for Information Science & Technology*, 52(3), 226-234.
- Tombros, A, Ruthven, I & Jose, J (2003). Searchers' criteria for assessing web pages. *26th ACM-SIGIR Conference*, ACM-Press, 385-386.
- Tombros, A, Ruthven, I & Jose, JM (2005). How users access web pages for information seeking.

Journal of the American Society for Information Science and Technology, 56(4), 327-344.

Vakkari, P (2001). Changes in search tactics and relevance judgments in preparing a research proposal: A summary of findings of a longitudinal study. *Information Retrieval*, 4(3/4), 295-310.

Vorhees, EM (2001). Evaluation by highly relevant documents. In: *24th ACM-SIGIR Conference*. ACM Press, 74-82.

Voorhees, EM (1998). Variations in relevance judgments and the measurement of retrieval effectiveness. In: Croft, WB. et al. (Eds.), *21st ACM-SIGIR Conference*. ACM Press, 315-323.

Wang, AP, de Vries, AP & Reinders, MJT (2006). In Lalmas, M., Tombros, A. (eds), *Proceedings of the Annual European Conference on Information Retrieval (ECIR)*, 37-48.

White, RW (2006). Using searcher simulations to redesign a polyrepresentative implicit feedback interface. *Information Processing and Management*, 42(5), 1185-1202.

White, R, Ruthven, I, José, JM & van Rijsbergen, CJ (2005). Evaluating implicit feedback models using searcher simulations. *ACM Transactions on Information Systems (ACM TOIS)*, 23(3), 225-361.

Willett, P (1988). Recent trends in hierarchic document clustering: A critic review. *Information Processing and Management*, 24(5), 577-597.