

Menneske-datamaskine interaktion i informationssøgssystemer: evaluering af brugs- og læsevenlighed

Kasper Hornbæk, Datalogisk Institut, Københavns Universitet

Informationssøgssystemer bør støtte mennesker i at søge efter, læse og på anden måde håndtere information. Ud fra min baggrund i fagfeltet menneske-datamaskine interaktion argumenterer jeg for at brugs- og læsevenlighed er nyttige evalueringskriterier for informationssøgssystemer. Denne argumentation er i overensstemmelse med en stigende interesse i fagfeltet informationssøgning for brugergrænseflader og for menneske-datamaskine interaktion. For at illustrere nytten af at fokusere på brugs- og læsevenlighed giver jeg to eksempler på evaluering af informationssøgssystemer som jeg har været med til at udvikle. Målet med brugergrænsefladerne til disse informationssøgssystemer er at støtte brugs- og læsevenlighed. Eksemplerne illustrerer hvordan vi i menneske-datamaskine interaktion arbejder med evaluering. De viser også hvordan tætte studier af interaktionen mellem mennesker og datamaskiner kan bruges til at forstå gode og dårlige egenskaber ved informationssøgssystemer.

Først i artiklen vil jeg beskrive hvad menneske-datamaskine interaktion handler om og hvorfor fagfelterne menneske-datamaskine interaktion og informationssøgning kan inspirere hinanden. Afsnittet 'Brugsvenlighed og læsevenlighed' viser hvordan brugs- og læsevenlighed kan måles i forbindelse med evaluering. De to næste afsnit i artiklen 'Tematiske kort' og 'Visualisering af dokumenter' beskriver de to eksempler på evaluering af brugergrænseflades brugs- og læsevenlighed. Det sidste afsnit i artiklen opsummerer nytten af fokus på menneske-datamaskine interaktion ved evaluering af informationssøgssystemer.

Menneske-datamaskine interaktion og informationssøgning

I menneske-datamaskine interaktion arbejder vi med design, evaluering og implementering af interaktive edb-systemer. Særligt prøver vi at forstå de fænomener som omgiver og former menneskers brug af edb-systemer. Fremkomsten af den personlige computer og den stigende brug af datamater i mange menneskers arbejde og fritid – blandt andet gennem World Wide Web (WWW) – har gjort udvikling af velegnede brugergrænseflader og studier af menneske-datamaskine interaktion nødvendig. På den baggrund er menneske-datamaskine interaktion vokset som fagfelt og har nu årlige konferencer (for eksempel ACMs konference om *Computer-Human Interaction*) og flere solide introducerende bøger (for eksempel Shneiderman 1998).

I menneske-datamaskine interaktion har der fra 1980'erne været en udbredt interesse for informationssøgning. Der bliver i stadig stigende omfang udviklet innovative grænseflader til informationssøgning som fx støtter formulering af forespørgsler (Young & Shneiderman 1993), bearbejdning af søgeresultater (Hearst 1995) eller giver et overblik over dokumentssamlinger (Chen et al. 1998). Mit eget arbejde i menneske-datamaskine interaktion har været koncentreret om informationssøgning,

blandt andet fordi problemet er genkendeligt og komplekst, men også fordi jeg mener at mange informationssøgssystemer kan forbedres afgørende, fx ved brug af grafiske brugergrænseflader.

I feltet informationssøgning er tilsvarende en stigende interesse for brugergrænseflader og menneske-datamaskine interaktion. For eksempel har Marchionini & Komlodi (1998) vist at ordene 'user interface' i databaserne *Library and Information Science Abstracts* og *Information Science Abstracts* forekommer i dobbelt så mange resumeer af dokumenter i 1996 som i 1986. Jeg tror denne interesse skyldes tre ting:

- I informationssøgning er der en stigende kritik af de mål for søgeeffektivitet som hyppigst anvendes i evaluering af systemer til informationssøgning, *precision* og *recall*, se fx Hersh (1994). Det er blevet klart at disse evalueringskriterier kun i begrænset omfang er relateret til den egentlige nytte og brugbarhed af informationssøgssystemer. Marchionini (1992), p. 156, har udtrykt det således:
'End users want to solve their information problems; they want to do their work or accomplish their goal. They want answers rather than pointers; they want document delivery rather than information retrieval. End users do not care about the design of the information system, the elegance of the retrieval engine, or the structure of the data unless these severely impede the accomplishments of their goals.'

Derfor er det vigtigt at finde evalueringsmetoder og evalueringskriterier som er relateret til hvor godt brugerne løser deres opgaver, ikke hvor godt systemet fremfinder dokumenter. Alternativer til *recall* og *precision*, såsom kvaliteten af de opgaver der løses ved brug af informationssøgssystemet og brugernes tilfredshed med systemet har i mange år været anvendt i menneske-datamaskine interaktion. Menneske-datamaskine interaktion har også i vid udstrækning brugt evalueringsmetoder som inddrager brugere i undersøgelser af informationssøgning.

- For det andet er det i stigende omfang klart at brugergrænsefladens udformning er kritisk afgørende for udfaldet af informationssøgning. Susan Dumais (1996) har påpeget at mens algoritmer til søgemaskiner ofte giver forbedringer fra 1-10% i målbare evalueringskriterier, så forbedrer veludformede brugergrænseflader informationssøgning med op til 25%. Kun systemer til relevans feedback kan opnå tilsvarende forbedringer. Denne udtalelse er særlig bemærkelsesværdi fordi Dumais i mange år har forsket i netop algoritmer til informationssøgssystemer. Veludformede brugergrænseflader og solide evalueringer af disse er derfor vigtigt.
- Endelig foretages stadig mere søgning af brugere uden særlige kundskaber om søgemaskiner, forespørgselsprog og katalogiseringsprincipper, fx på WWW. Dette stiller krav til informationssøgssystemer om at støtte novicer i at søge og på at gøre det nemt at lave forespørgsler og forstå resultaterne af søgninger. Menneske-datamaskine interaktion tilbyder teknikker som forbedre disse dele af en brugergrænseflade.

Brugs- og læsevenlighed

For menneske-datamaskine interaktion er brugs- og læsevenlighed afgørende ved design og evaluering af informationssøgssystemer.

Brugsvenlighed er det væsentligste begreb fra menneske-datamaskine interaktion i forbindelse med evaluering af informationssøgessystemer. Der findes en række forskellige forslag om hvad brugsvenlighed omfatter og hvordan man måler brugsvenlighed, se fx Shneiderman (1998). Disse forslag omfatter den tid det tager at lære at bruge systemet, hvor nemt det er at huske hvordan systemet bruges, den tid det tager at løse bestemte opgaver, nytten af systemet og brugernes tilfredshed med systemet. ISO har foreslået en standard for brugsvenlighed (ISO 1998) som omfatter tre aspekter.

- Effektivitet, det vil sige hvor præcist og komplet brugeren når sine mål. Effektivitet kan måles fx som kvaliteten af løsning på de opgaver som er blevet støttet af informationssøgessystemet.
- Produktivitet, som er forholdet mellem effektivitet og de ressourcer som brugeren anvender for at nå sine mål. Produktivitet vil ofte være udtrykt som fx den tid det tager at løse en opgave eller den mentale belastning som det medfører at løse en opgave med systemet.
- Tilfredshed, som er brugerens holdning til og komfort ved brugen af systemet. Brugernes tilfredshed vil ofte blive målt ved åbne spørgsmål som stilles efter brug af systemet eller med spørgeskemaer dedikeret til måling af tilfredshed, fx Questionnaire for User Interaction Satisfaction (Chin et al. 1988).

Sammen med Erik Frøkjær og Morten Hertzum har jeg argumenteret for at alle tre aspekter af brugsvenlighed skal måles i forbindelse med evaluering (Frøkjær et al. 2000). Da man ikke kan antage at de tre aspekter hænger sammen bør man måle dem alle.

Systemer til informationssøgning bør tillige støtte *læsevenlighed*. Dette er en konsekvens af at dokumenter, web-sider og e-post i stigende grad læses direkte fra skærme. For eksempel har Sellen & Harper (1997) beskrevet brugen af papir og elektroniske dokumenter hos analytikere i International Monetary Fund. Sellen & Harper anslår at 14% af tiden analytikerne arbejdede med dokumenter bruger de dokumenter som de læser fra skærmen. Analytikerne brugte en kombination af skærlæsning og papirlæsning 35% af tiden. Et studie af brugen af WWW (Byrne et al. 1999) viste at brugere benytter dobbelt så lang tid på at bruge den information de finder som på søgning, browsing eller andre aktiviteter. Læsning fra skærmen udgør den vigtigste måde at bruge information på. I menneske-datamaskine interaktion, i pædagogisk orienterede undersøgelser af læsning og i teknikker til effektiv udformning af dokumenter er der forsket intenst i hvordan en brugergrænseflade kan støtte læsevenlighed (Dillon 1994; Schriver 1997). Denne litteratur beskriver teknikker til at forbedre navigation i dokumenter, til at gøre læsning fra skærmen mere behagelig og til at hjælpe brugere med bedre at huske en tekst. Det er muligt at forbedre læsevenligheden uden at ændre dokumenternes indhold, fx ved at sikre at skrifttyperne er læselige, at linierne i dokumentet ikke er for lange, eller ved at tilbyde et overblik over dokumenter. Når brugere læser direkte fra skærmen bliver løsningen af det problem som ledte til informationssøgning afhængig af hvor godt systemet støtter læsning. Læsevenlighed er derved en forudsætning for brugsvenlighed og kan måles indirekte ved at se på brugsvenligheden af et system. Læsevenlighed kan også evalueres direkte ved at se på hvordan brugerne navigerer i dokumenter, hvordan de læser dokumenter, hvor meget de lærer af at læse teksten, samt ved at undersøge deres tilfredshed med læsningen.

Desværre er evaluering af læsevenlighed endnu ikke udbredt i evaluering af systemer til informationssøgning (undtagelser findes, fx Kengeri et al. 1999). Dette undrer mig eftersom forskning i informationssøgning i 50 år har fokuseret på at støtte fremfindning af dokumenter uden dog at støtte brugere i bekvemt og effektivt at læse disse dokumenter fra en skærm.

I de følgende to afsnit giver jeg eksempler på hvordan evalueringer af systemer til informationssøgning og -håndtering kan foregå ud fra et fokus på brugsvenlighed og læsevenlighed.

Tematiske kort

Tematiske kort er blevet foreslået som en måde at forbedre informationssøgning på. Et tematisk kort viser et grafisk overblik over en samling af dokumenter sammen med indikationer af temaer i dokumentetsamlingen, se figur 2. Forhåbningen til tematiske kort er at de forbedrer brugsvenligheden i forhold til ikke-grafiske systemer. Det tematiske kort antages at være nemmere bruge til at overskue en dokumentetsamling, mere behageligt at søge på og hurtigere fordi brugere kan søge ved hjælp af den visuelle repræsentation af dokumenter og temaer. For at undersøge om disse forhåbninger kunne indfries sammenlignede vi et tekstbaseret informationssøgningssystem med et tematisk kort. Både det tekstbaserede system og det tematiske kort tillod forsøgspersonerne at formulere forespørgsler og at se fuldttekst af dokumenterne i samlingen – den visuelle repræsentation af dokumenter og temaer er den eneste forskel mellem systemerne og derfor den eneste kilde til forskel i brugsvenlighed. Undersøgelsen er yderligere beskrevet i Hornbæk & Frøkjær (1999).

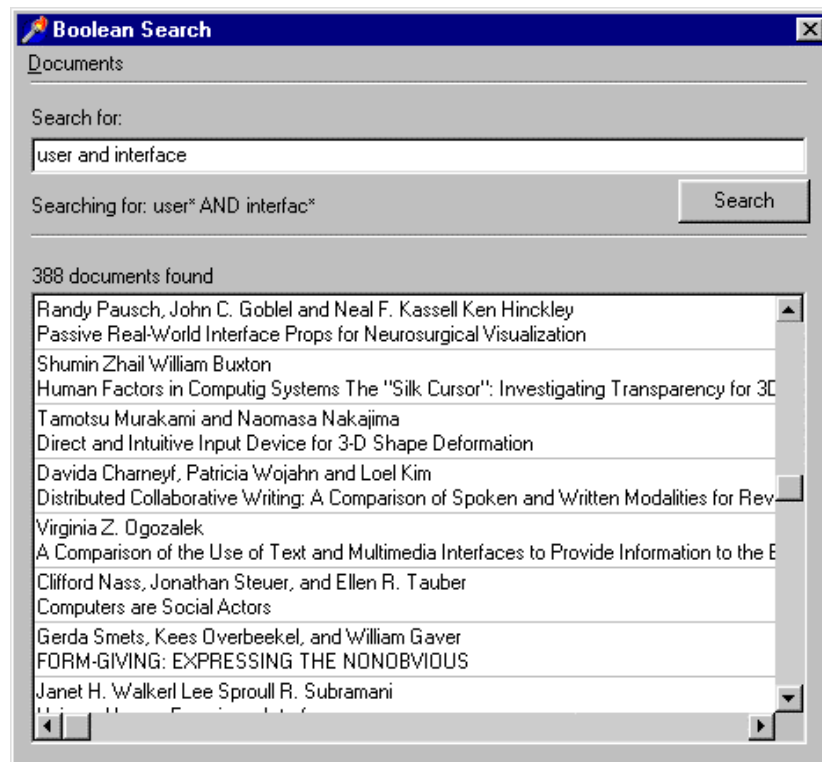
Undersøgelsens opbygning

I undersøgelsen løste seks forsøgsdeltagere hver otte opgaver, fire opgaver med hver af de to grænseflader. Halvdelen af opgaverne beskrev ét dokument som forsøgspersonerne skulle finde i dokumentetsamlingen, fx 'Find artiklen af Rudolf Darken om hvordan man navigerer virtuelle verdener'. Den anden halvdel af opgaverne skulle besvares med en gruppe af dokumenter, fx 'Forestil dig at du skal holde et foredrag om brug af computere i undervisning. Hvad findes der i systemet om dette emne?'. Dokumentetsamlingen som forsøgspersonerne søgte i indeholdt omkring 500 fuldttekstartikler fra konferencer om menneske-datamaskine interaktion. Undersøgelsen blev gennemført med forsøgspersoner som skrev speciale i datalogi og kendte til menneske-datamaskine interaktion. Alle forsøgspersonerne var fortrolige med boolske operatorer. Under eksperimentet blev forespørgsler, læsning af dokumenternes fuldttekst og brug af det tematiske kort automatisk registreret. Det var derfor muligt efter forsøget at rekonstruere søgningerne. Forsøgspersonerne blev opfordret til at tænke højt mens de løste opgaverne. Tænke-højt teknikker er almindeligt anvendt i HCI og går ud på at få forsøgspersoner til hele tiden at fortælle hvad de tænker, se Helms Jørgensen (1990). Forsøgspersonerne ytringer blev optaget på bånd.

Tekstbaseret informationssøgningssystem og tematisk kort

Det tekstbaserede informationssøgningssystem er vist i figur 1. Brugeren kan formulere forespørgsler med boolske operatorer (øverst på figuren). Resultatet af forespørgsler vises

som en liste af titler og forfattere (nederst på figuren). Fuldteksten af dokumenter i listen kan ses ved at dobbeltklikke med musen på en titel.



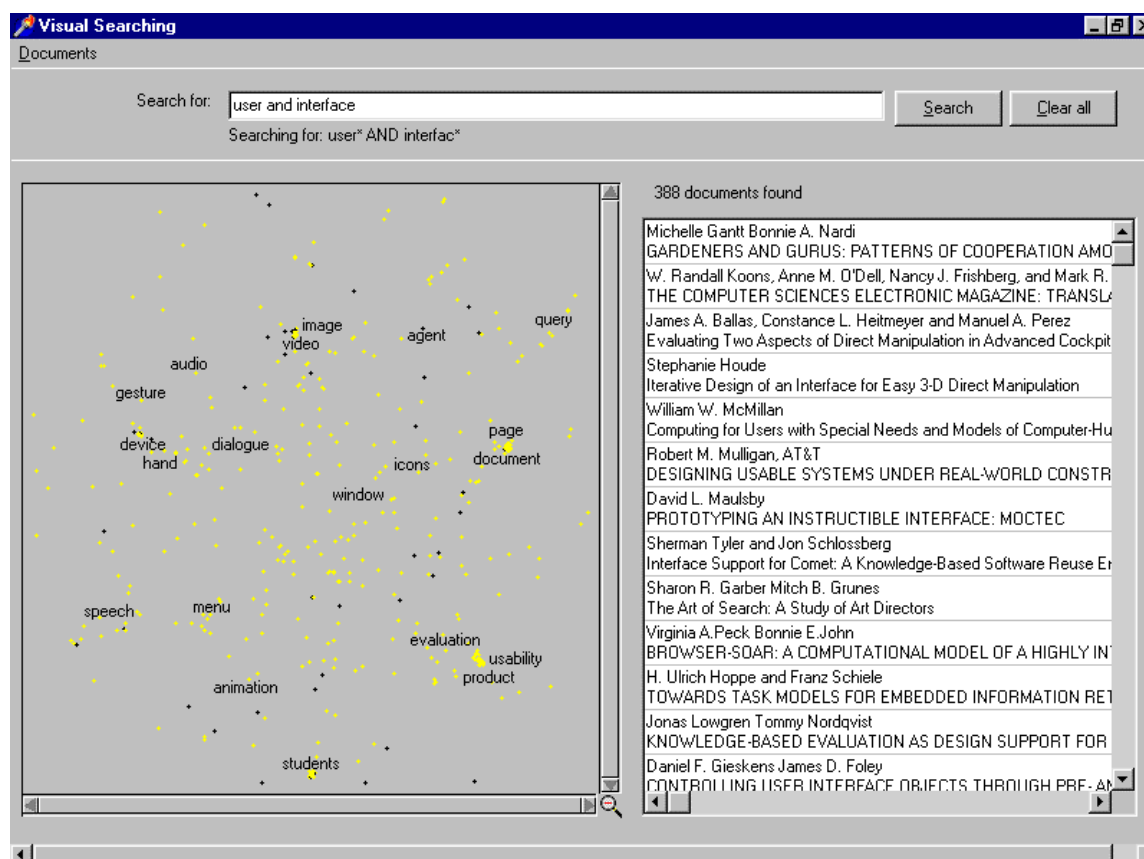
Figur 1 – Tekstbaseret informationssøgesystem.

Det tematiske kort er vist i figur 2. Feltet øverst på skærmen til indtastninger af forespørgsler og listen til højre over fremfundne dokumenter svarer til den øverste og nederste del af det tekstbaserede informationssøgesystem. Til venstre findes selve det tematiske kort med dokumenter og ord som beskriver temaer på kortet. Dokumenter er vist som sorte og lyse prikker (gule på kortet). Dokumenterne der er blevet fremfundet ved søgning efter ordene 'user' og 'interface', er markeret med en lys prik på kortet. De sorte prikker er dokumenter som ikke indeholder ordene 'user' og 'interface'.

Det tematiske kort er konstrueret ved brug af multidimensionel skalering (Borg & Groenen 1997). Algoritmen til multidimensionel skalering placerer dokumenterne i dokumentsamlingen i to dimensioner ved brug af cosinus målet for lighed mellem dokumenter – et almindeligt benyttet lighedsmål i søgemaskiner (Salton & McGill 1983). Placeringen af dokumenter på kortet sigter mod at minimere forskellen mellem cosinus ligheder mellem dokumenter og afstande mellem dokumenterne på kortet. Udover dokumenterne viser det tematiske kort også ord som beskriver temaerne i dokumenterne. De ord som bliver placeret på kortet er de ord i dokumentsamlingen med højest diskriminationsværdi (Salton & McGill 1983). Intuitivt er ord med høj diskriminationsværdi hyppigt forekommende i nogle dokumenter og sjældent i andre, for eksempel ordet 'evaluation' i den brugte dokumentsamling. Disse ord blev placeret på

massemidtpunktet mellem dokumenter som de forekommer i og beskriver derved temaerne i det område af kortet de er placeret i.

Forespørgsler, listen over dokumenter og det tematiske kort er knyttet sammen på flere måder. Ud over at alle dokumenter på kortet som er fremfundet som svar på en forespørgsel farves gule, indikeres forespørgselstermer også visuelt på kortet. Når brugeren vælger et dokument på listen med fremfundne dokumenter, markeres dette dokumentets position på kortet. Dokumenternes titler og forfattere vises når musen holdes over et dokument på kortet, ligesom fuldt tekst kan bringes frem fra kortet ved hjælp af en pop-up menu. Det er muligt at zoome ind på en del af kortet hvorved der fremkommer flere karakteriserende termer.



Figur 2 – Tematisk kort

Forskelle i brugsvenlighed

Brugsvenligheden af de to grænseflader blev målt som antallet af dokumenter forsøgspersonerne fandt relevante, deres tidsforbrug og deres tilfredshed med systemet. Dertil har vi også en række kvalitative observationer som belyser hvordan forsøgspersonerne bruger kortet.

Vi fandt ingen statistisk signifikant forskel mellem de to grænseflader i antallet af dokumenter som blev markeret som relevante af forsøgspersonerne og som vi også fandt relevante (dvs. interaktiv *recall*). Antallet af dokumenter som mere end en person fandt relevant var heller ikke væsentligt forskelligt mellem grænsefladerne. Tilgængæld fandt vi

en overraskende forskel i tidsforbrug mellem de to grænseflader. Forsøgspersonerne brugte 14.2 minutter per opgave med den visuelle grænseflade og blot cirka 10.8 minutter med den boolske grænseflade. Vores egne og gængse hypoteser om nytten af tematiske kort kan derved ikke indfries i dette studie. Efter at have løst opgaverne blev forsøgspersonerne interviewet om deres tilfredshed med de to grænseflader. Fire ud af de seks forsøgspersoner foretrak at bruge det tematiske kort: de synes at grafikken var behagelig, de kunne lide det overblik man fik fra det tematiske kort og synes at de blev inspireret til at finde søgeord ved at kigge på kortet. En af forsøgspersonerne sagde:

'Jeg kunne bedre lide det grafiske, det var lidt sjovere på en eller anden måde. Det er nok det bedste ved [det tematiske kort]'

og en anden sagde:

'Dels havde man ordene [på kortet] og så kunne man også se hvor mange [dokumenter] man havde ramt sådan nogenlunde, så man kunne hurtigere danne sig et overblik over hvordan ens søgning gik, om man havde ramt mange eller få, også hvordan de lå placeret i forhold til hinanden, om de var tæt på eller spredt.'

I interviewet efter forsøget nævnte halvdelen af forsøgspersonerne at de til tider havde problemer med at forstå kortet, fx var forholdet mellem ord og dokumenter på kortet uklart for nogle af forsøgspersonerne. En af dem nævnte:

'Jeg undrer mig lidt over de kategorier [ord på kortet] som er vist, det er lidt ... nogle af dem er hovedtemaer i datalogien ligesom 'evaluation' og 'usability' som man kan forholde sig til men sådan noget som 'hand' ... det kan betyde hvad som helst'

og en anden:

'Det var lidt uklart hvad det [kortet] egentlig betød. Jeg fik ikke nogen fornemmelse af en struktur eller gruppering. Lidt abstrakt hvad det betyder med nogle prikker rundt omkring.'

Flere forsøgspersoner udtrykte også overraskelse når de så på dokumenter ved siden af hinanden på kortet og ikke kunne afgøre hvad dokumenterne havde tilfældes. En forsøgsperson kommenterede 'Man håber at når dokumenterne er tæt på hinanden så handler de om det samme'. Da jeg spurgte forsøgspersonen hvor tit dokumenterne handlede om det samme sagde han 'Måske halvdelen af gangene.' Tilfredsheden med det tematiske kort er derved blandet, med positive kommentarer om oplevelsen af at søge men også væsentlige problemer med kortets forståelighed og gennemskuelighed.

De kvalitative observationer om brugen af kortet vedrører forsøgspersonernes inspiration til søgeord og hvordan forsøgspersonerne brugte kortet til at søge på. Ved at analysere forsøgspersonernes ytringer fra eksperimentet fandt vi fire kilder som inspirerede forsøgspersonernes valg af søgeord: 1) inspiration fra teksten som beskriver opgaven, 2) inspiration fra associationer eller anden inspiration som ikke kan detekteres i ytringerne, 3) inspiration fra dokumenternes fuldtekst eller titler og 4) inspiration fra ord på det tematiske kort. Med denne klassifikation viser det sig at cirka halvdelen af søgeordene kommer fra beskrivelsen af opgaven, cirka en tredjedel fra associationer og cirka ti procent fra dokumenters fuldtekst henholdsvis fra det tematiske kort. Typisk sker inspirationen fra det tematiske kort til søgeord ved at forsøgspersonen ser et ord på kortet og derefter bruger det i en forespørgsel.

Forsøgspersonerne foretrak at søge på kortet frem for at skimme listen af titler og forfattere: 1/3 af alle opgaver blev startet på kortet, af de resterende opgaver begyndtes

resten med en forespørgsel men derefter begyndte forsøgspersonerne at søge på kortet. Når forsøgspersonerne søgte på kortet var der tre typiske måder at søge på. For det første fokuserede nogle forsøgspersoner på bestemte ord eller områder på kortet. For eksempel i en opgave om lyd i brugergrænseflader bemærkede en forsøgsperson:

‘Ja, grænseflader der anvender lyd, så var det jo at der lå, hvad var det jeg fandt der hed noget her ’audio’, ’speech’ på kortet, netop fordi det sådan er et søgeord der, tror jeg jeg vil gå ind og kigge om der er noget [på kortet]’

Nogle forsøgspersoner fokuserede også på et bestemt område på kortet, fx opdagede en person et område med mange dokumenter fremfundet i relation til en forespørgsel og sagde:

“Jeg prøver lige at se på den sky der over [på kortet], for at se hvorfor den er placeret derovre”.

For det andet var der opgaver hvor forsøgspersonerne brugte et dokument position på kortet til at vurdere om dokumentet var relevant. Forsøgspersonerne brugte også et dokument position til at finde andre relevante dokumenter.:

” hmm, ja det er jo, altså jeg ville jo umiddelbart sige at det er sikkert også relevant fordi den ligger i nærheden af de andre [fremfundne dokumenter]”.

For det tredje inspirerede det tematiske kort nogle forsøgspersoner til at browse målløst. Vi observerede primært dette i søgninger efter velspecificerede dokumenter, hvor nogle forsøgspersoner – når de ikke kunne finde et dokument som besvarede opgaven – så mange gange på de samme områder og dokumenter.

Konklusion på undersøgelsen af tematiske kort

Undersøgelsens hypotese om at en visuel grænseflade forbedrede brugsvenligheden af informationssøgning blev ikke entydigt bekræftet. Det tematiske kort var langsomme end den boolske grænseflade. Omvendt viser interviewene med forsøgspersonerne at de kunne lide at søge på kortet og analysen af tænke-højt protokollen viser at de fandt nyttige søgeord på kortet. Interaktionen med kortet er omfattende og er blandt andet rettet mod at prøve at afgøre dokumenters relevans ud fra deres position på kortet og mod at finde interessante ord og områder på kortet. Hvorfor er brugsvenligheden så ikke højere for det tematiske kort? En forklaring er at forsøgspersonerne taber fokus på deres søgeopgave, fordi de skifter meget mellem forespørgsler og at skimme kortet og fordi de til tider distraheres af det tematiske kort. Desuden er forståeligheden af det automatisk producerede kort problemfyldt.

En klar forbedring af det tematiske kort ville være bedre algoritmer til at vælge karakteriserende termer som placeres på kortet, fx kunne man give mere kontekst end et enkelt ord ved at bruge kombinationer af ord eller sætninger. Dertil bør forståeligheden af kortet forbedres generelt, fx ved at gøre det eksplicit hvilke temaer dokumenter tæt på hinanden har tilfælles. Desuden er det svært at foretage en systematisk afsøgning af fremfundne dokumenter på kortet: flere forsøgspersoner foreslog at en mere håndterbar, lineær struktur ville være et nyttigt supplement til kortets associative struktur.

Visualisering af dokumenter

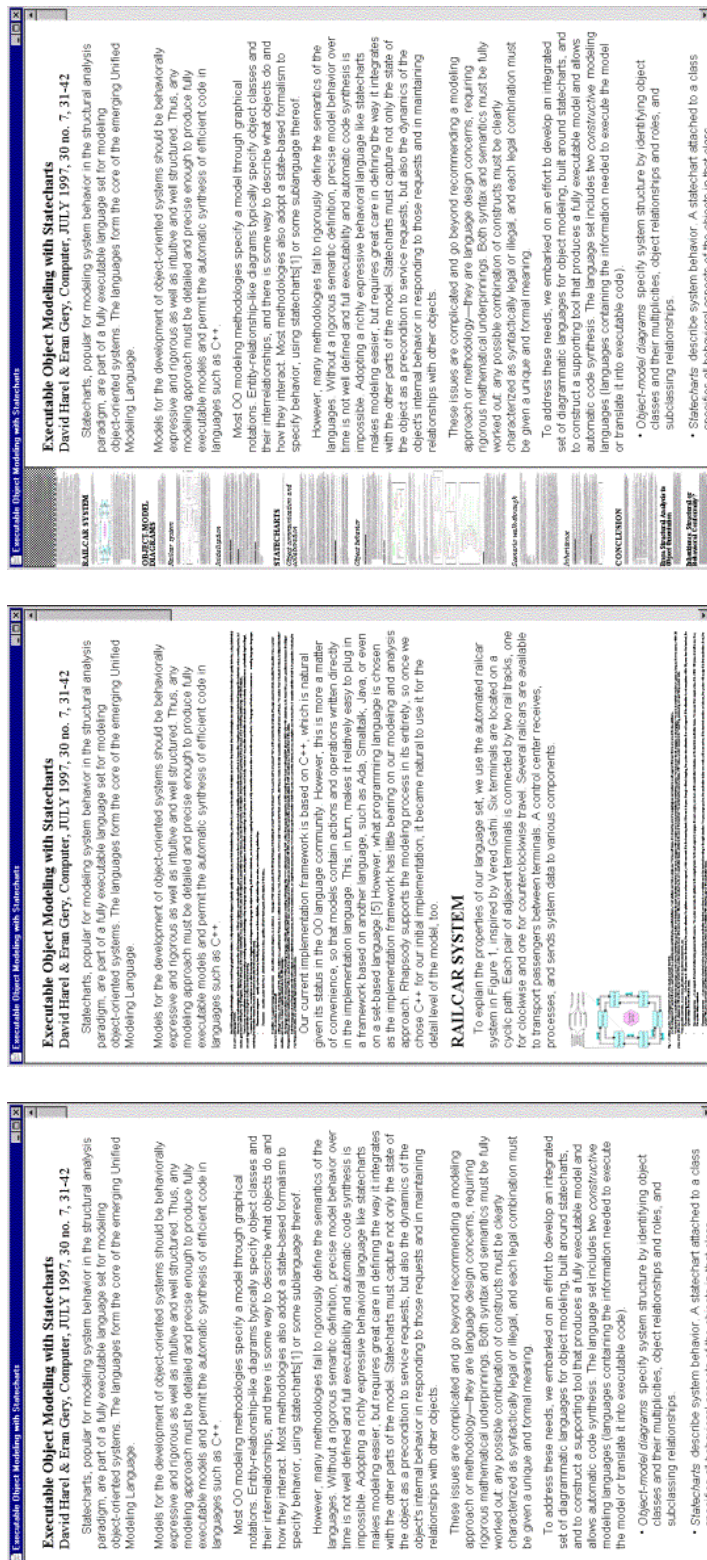
Litteraturen om menneske-datamaskine foreslår adskillige brugergrænseflader som visualiserer dokumenter, se fx Card et al. (1999). Intentionen med disse brugergrænseflader er at støtte læsning og navigation i dokumenter ved at vise dokumenternes struktur eller ved visuelt at fremhæve vigtige dele af dokumentet og nedtone andre dele. Brugs- og læsevenligheden af de foreslåede grænseflader er dog kun i et par tilfælde blevet undersøgt. Disse undersøgelser har ikke kunne vise at brugen af visualisering giver højere brugsvenlighed. Derfor gennemførte vi et eksperiment hvor 20 forsøgspersoner læste artikler om objekt-orienteret system udvikling med en lineær, en fiskeøj og en overbliksgrenseflade (se figur 3). Vi ønskede at undersøge forskelle i brugsvenlighed mellem brugergrænsefladerne. Til de formål så vi på karakterer givet til forsøgspersonernes svar på opgaver vi gav dem, på forsøgspersonernes tilfredshed og på interaktionen med grænsefladerne. Desuden ville vi gerne undersøge læsevenlighed og beskrive hvordan forskellige brugergrænseflader påvirker læseprocessen. Eksperimentet er yderligere beskrevet i (Hornbæk & Frøkjær 2001).

Brugergrænseflader

Figur 3 viser skærbilleder af de tre brugergrænseflader som vi undersøgte i eksperimentet. I alle grænseflader kan musen og tastaturet bruges til at navigere med. Den lineære grænseflade (nederst i figur 3) viser dokumenter på samme måde som de fleste tekstbehandlingsprogrammer og WWW-browsere. Hensigten med denne grænseflade er at etablere et sammenligningsgrundlag for de to andre grænseflader.

Fiskeøjegrænsefladen (midterst i figur 3) behandler nogle dele af et dokument som mere vigtige end andre. Fiskeøjegrænseflader fungerer på samme måde som et fiskeøjobjektiv til et kamera, ved at forstørre de centrale/vigtige dele af motivet mens de perifære dele af motivet vises uforholdsmæssigt små. I brugergrænsefladen brugt i eksperimentet er de vigtige dele af et dokument er altid læsbare. De øvrige dele af dokumentet er formindsket, men kan udvides til læselig størrelse ved at klikke på dem med musen. De vigtige dele af en artikel bestemmes på to måder. For det første viser forskning i automatisk sammenfatning af dokumenter at de første og sidste dele af et afsnit er gode indikatorer for afsnittets indhold (Bradow et al. 1995). Vi brugte denne ide på alle sektioner i dokumentet, så de midterste dele af alle afsnit var formindskede. For det andet lægger læsere af videnskabelige dokumenter særligt mærke til resumeet, overskrifter og så videre (Dillon 1994; Bishop 1999). Derfor kan man i fiskeøjegrænsefladen altid læse disse dele af dokumenterne. Hensigten med fiskeøjegrænsefladen er at reducere den tid der bruges på at navigere i dokumentet ved at gøre dokumentet kortere – i gennemsnit er dokumenterne 25% kortere i fiskeøjegrænsefladen. Desuden ville vi støtte forsøgspersonerne i en overbliksoorienteret læsestrategi, hvor de først ser de vigtigste afsnit og senere kan læse detaljerne.

I overbliksgrensefladen (øverst i figur 3) er dokumentet vist i højre del af grænsefladen – detaljeområdet – på samme måde som i den lineære grænseflade. Dertil er knyttet et overblik over dokumentet – overbliksområdet – som viser en kompakt version af dokumentet hvor overskrifter på første og andet niveau i dokumentet er synlige. Overbliksområdet indikerer ved en mørk firkant hvilken del af dokumentet der er vist i detaljeområdet. Det er muligt at bruge overbliksområdet til at navigere i artiklen, fx ved



Figur 3 – Grænseflader som visualiserer dokumenter

at klikke på en bestemt del af dokumentet som man ønsker at se. Hensigten med overbliksområdet er at støtte navigation ved hjælp af overbliksområdet, samt at give læserne en fornemmelse af dokumentets indhold og struktur

Eksperiment

De 20 forsøgspersoner løste to slags opgaver: essay-opgaver og spørgsmålsopgaver. I essay-opgaverne skulle forsøgspersonerne læse et helt dokument. Derefter skulle de skrive et essay som fyldte en side og beskrev indholdet af dokumentet. Desuden skulle de skrive en sides kommentarer om dokumentet. Efter at have skrevet essays og kommentarer skulle forsøgspersonerne besvare seks spørgsmål som de ikke kendte ordlyden af mens de læste dokumentet. Intentionen med disse 'incidental learning' spørgsmål var at afdække dybden af forsøgspersonernes læsning af dokumenterne. I spørgsmålsopgaver skulle forsøgspersonerne besvare seks spørgsmål, et af gangen. Det var meningen at forsøgspersonerne skulle lede i dokumenterne efter svaret. Dokumenterne som forsøgspersonerne læste var tidsskriftsartikler om objektorienteret systemudvikling på mellem 8 og 14 sider.

Alle forsøgspersonerne gennemførte først en træningsopgave, som i gennemsnit varede 35 minutter. Derefter løste forsøgspersonerne en essay-opgave og en spørgsmålsopgave. Dagen efter kom forsøgspersonerne tilbage og løste yderligere fire opgaver. Hvert par af en spørgsmålsopgave og en essayopgave blev løst med en af de tre grænseflader. I gennemsnit læste hver forsøgsperson over fem timer.

Forskelle i brugsvenlighed

For at vurdere brugergrænsefladens effektivitet gav vi svarene på opgaverne en karakter mellem 1 og 5. Karakteren 1 blev givet for det helt forkerte eller meget mangelfulde svar; karakteren 5 blev givet til den korrekte og meget velunderbyggede besvarelse. Mellem de tre grænseflader fandt vi en statistisk signifikant forskel i den karakter som forsøgspersonerne fik for besvarelsen af essay-opgaver. Med overbliksgrensefladen får forsøgspersonerne i gennemsnit en halv karakter mere end med den lineære og fiskeøjegrensefladen. Vi fandt også en signifikant forskel mellem grænseflader i de 'incidental learning' spørgsmål som forsøgspersonerne fik efter at have skrevet essayet. Med fiskeøjegrensefladen besvarer forsøgspersonerne mellem 0.8 og 1.2 færre spørgsmål korrekt end med den lineære og overbliksgrensefladen. Forsøgspersonernes tidsforbrug ved læsning af dokumenter som de skal skrive essays om er forskelligt: Fiskeøjegrensefladen er 16% hurtigere end overbliksgrensefladen og den lineære grænseflade.

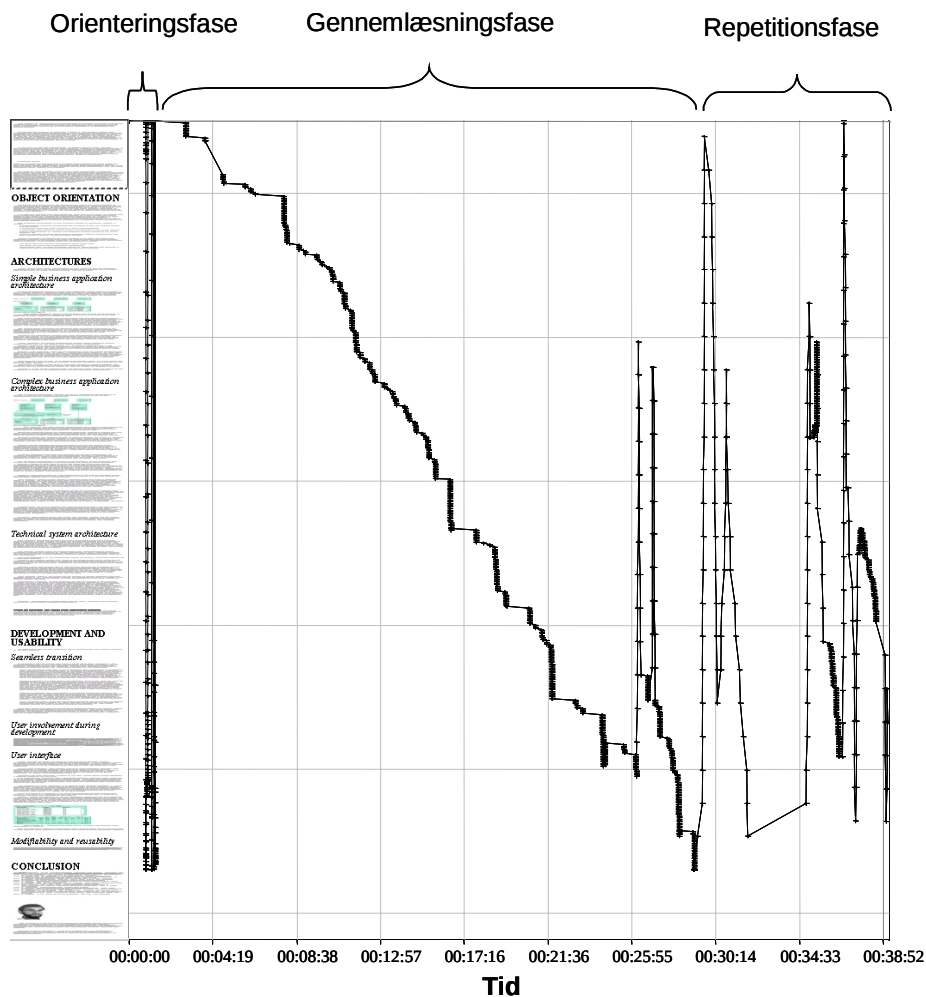
Vi fandt ikke nogen forskel i karakter mellem grænsefladerne for spørgsmålsopgaver, men overbliksgrensefladen er omkring 20% langsommere end den lineære grænseflade. Der er ikke nogle signifikante forskelle mellem den lineære grænseflade og fiskeøjegrensefladen.

Nitten af forsøgspersonerne foretrak at bruge overbliksgrensefladen, en person foretrak den lineære grænseflade. Ti af forsøgspersonerne nævnte overblikket over dokumentstrukturen som årsag til deres præference. Seks personer mente at overbliksgrensefladen støttede navigation og at den gjorde det nemme at finde rundt i dokumentet.

Forskelle i læsevenlighed

Ovenstående data belyser væsentlige forskelle i brugervenlighed mellem grænsefladerne. Disse data fortæller os dog ikke hvordan forsøgspersonerne læser dokumenterne eller hvordan de tre grænseflader støtter forsøgspersonerne i at navigere i dokumenterne. Under eksperimentet blev det automatisk registreret hvilke dele af artiklen forsøgspersonerne kunne se på et givent tidspunkt og hvordan de brugte mus og tastatur til at navigere igennem dokumentet. Derfor kan vi rekonstruere hvilke dele af dokumenterne forsøgspersonerne har kigget på, hvor lang tid de har kigget på forskellige dele af dokumentet, samt rækkefølgen de kiggede på dokumentets dele. Fra disse data kan vi vurdere hvordan de tre grænseflader påvirker læseprocessen.

For essayopgaverne finder vi at forsøgspersonernes læsning af dokumenterne foregår i tre forskellige faser: en orienteringsfase, en lineær gennemlæsning og en repetitionsfase. I orienteringsfasen navigerer forsøgspersonerne gennem dokumentet, mens de særligt ser på artiklens abstrakt, introduktion og konklusion. I den lineære gennemlæsningsfase læser forsøgspersonerne gennem dokumentet fra begyndelsen til slutningen, ofte med hop fremad i dokumentet og regressioner til tidligere læste dele. I repetitionsfasen synes forsøgspersonerne at genlæse det de opfatter som de vigtigste dele af artiklen. Figur 4 viser et eksempel på hvordan de tre faser kan se ud for en forsøgsperson. Den vandrette akse i figuren viser den tid som er forløbet siden starten af opgaven. Den lodrette akse viser position i dokumentet. Linierne i figuren viser hvilken del af artiklen forsøgspersonen har kunnet se på et bestemt tidspunkt. Øverst i figuren er vist varigheden af de tre faser. Det interessante ved disse faser er at længden af dem varierer mellem de tre grænseflader. I fiskeøjegrænsefladen bruger forsøgspersonerne i gennemsnit ni minutter på at orientere sig indledningsvis. Dette er mere end dobbelt så meget som forsøgspersoner der bruger de to andre grænseflader. Til gengæld bruger forsøgspersonerne cirka en tredje del mindre tid med fiskeøjegrænsefladen i den lineære gennemlæsningsfase end med de andre grænseflader. Derved synes fiskeøjegrænsefladen at invitere til en mere overbliksorienteret læsestil – hvilket muliggør at forsøgspersonerne med denne grænseflade løser essayopgaverne hurtigere. Fiskeøjegrænsefladen støtter dermed visse typer opgaver, fx relevansbedømmelse af dokumenter eller tidskritiske tilegnelsesopgaver, men støtter ikke lige så effektivt som de to andre grænseflader læsning for at opnå en fuld forståelse af dokumenter. [ToDo: læsemønstre]



Figur 4 – Læsemønstre for en essayopgave.

Læsevenligheden i forbindelse med spørgsmålsopgaverne kan undersøges på tilsvarende måde. I spørgsmålsopgaverne kan svarene på spørgsmål findes på bestemte positioner, såkaldte svarpositioner. Fra læsemønstrene kan vi afgøre hvornår en svarposition første gang ses og hvad forsøgspersonerne gør når de har fundet en svarposition. En tæt undersøgelse af disse data viser to ting om læsevenlighed i de tre grænseflader. For det første når forsøgspersonerne cirka lige hurtigt frem til en svarposition med de tre grænseflader. De betyder at forskellen i tidsforbrug ikke kan skyldes at en af grænsefladerne ikke støtter at finde et svar hurtigt. Tilgængeligt er det klart fra vores data at forsøgspersonerne med overbliksgrensefladen cirka 40% hyppigere navigerer væk fra et mål efter de har set på svaret i forhold til den lineære grænseflade. Derved finder forsøgspersoner som bruger overbliksgrensefladen 10% flere mål i de artikler hvor svaret kan findes flere steder i artiklen. Overbliksgrensefladen indbyder derved til udforskning af artikler og tilbyder åbenbart bekvemme måder at gøre dette på. Omvendt er nogle læseprocesser måske skrøbelige i den forstand at overbliksgrensefladen vil distrahere læsere og lede dem til at miste fokus på den opgave

de er ved at løse. De højere tidsforbrug i spørgsmålsopgaver kan måske skyldes at overblikket i nogle situationer tiltrækker sig forsøgspersonernes opmærksomhed og derved får dem til yderligere at udforske dokumentet, selvom de allerede har fundet et anvendeligt svar på opgaver.

En anden interessant observation er at forsøgspersoner med overbliksgrensefladen mere hyppigt vender tilbage til tidligere sete svarposition end med de andre grenseflader. I stort omfang bruger de overbliksområdet til at gøre dette. Måske forsøgspersonerne erindrer hvilket område af dokumentet de fandt bestemte svar. Fra forskning i papirdokumenter ved vi at læsere, uden at gøre sig det bevidst, erindrer hvor på siden information stod (Rothkopf 1971). Noget lignende er normalt sværere at gøre i elektroniske dokumenter, særligt når teksten er kontinuert. Overbliksgrensefladen støtter på den måde læsning ved at støtte læseren i at forbinde en position med det der læses.

Konklusion på undersøgelsen af visualisering af dokumenter

Ud fra vores eksperiment mener vi at overbliksgrensefladen støtter forsøgspersonerne bedre i at læse end de to andre grenseflader. Forsøgspersonerne fik højere karakterer for deres essays. De foretrak også at læse ved hjælp af overbliksgrensefladen og hævdede at de fik et bedre overblik over dokumentet. Desuden støttes forsøgspersonerne i at huske tekstfragmenters position i dokumentet. Derfor mener vi at langt flere informationssøgesystemer burde anvende grenseflader som overbliksgrensefladen og derved øge systemets løsevenlighed.

Fiskeøjegrensefladen støtter forsøgspersonerne i at læse hurtigt og støtter en overbliksorienteret læsestil. Nogle forsøgspersoner var dog utrygge ved om vigtige afsnit af teksten var blevet formindsket. Desuden viser vores 'incidental learning' spørgsmål at forsøgspersonerne ikke får samme dybe forståelse af de dokumenter de læser som med de andre grenseflader.

Vores eksperimentet viser nytten at undersøge læsning tæt. Først da vi kiggede på løsevenligheden af systemerne kunne vi forklare hvorfor vi observerede forskelle i brugsvenlighed.

Konklusion

Fra mit arbejde i menneske-datamaskine interaktion har jeg i denne artikel forsøgt at vise hvordan brugsvenlighed og læsevenlighed kan indgå i evaluering af informationssøgesystemer. Brugsvenlighed giver fokus på hvor effektive, produktive og tilfredse brugere er når de søger information. Læsevenlighed vedrører tilegnelse af dokumenter og støtte til navigation i dokumenterne. Tilsammen giver de to begreber fokus på interaktiv informationssøgning, på design af velfungerende brugergrænseflader og på at relatere evaluering til brugeres faktiske gavn af informationssøgesystemer.

Jeg illustrerede nytten af begreberne brugs- og læsevenlighed ved to eksempler. Evalueringen af det tematiske kort viser hvordan almindeligt forekommende antagelser om værdien af tematiske kort ikke viser sig som forskelle i brugsvenlighed. Til gengæld afdækkede evalueringen en modsætning mellem den tid forsøgspersonerne brugte på at søge og deres tilfredshed med systemerne. Kun ved at bruge en vifte af mål for brugsvenlighed kan vi nå til en nogenlunde forståelse af gode og dårlige egenskaber ved tematiske kort. Eksperimentet med visualisering af dokumenter viser at det er muligt at fokusere på læseprocessen ved design og evaluering af informationssøgesystemer. Derved kunne overbliksgrensefladen forbedre den karakter forsøgspersonerne opnåede med cirka 25% – en markant forbedring ud fra et fokus på læsevenlighed.

De to eksempler viser også nytten af at forstå forskelle i brugs- og læsevenlighed ved at se på hvordan brugere interagerer med informationssøgesystemer. Kun ved at undersøge hvordan det tematiske kort blev brugt og hvordan forsøgspersonerne interagerede med visualiseringer af dokumenter kan vi opstille forklaringer på hvorfor der er forskelle i brugs- og læsevenlighed. Eksemplerne har vist hvordan visuelle grænseflader kan lede brugeren til at udforske dokumenter og dokumentsamlinger unødigt – om dette er hensigtsmæssigt afhænger af hvilke opgaver informationssøgesystemet skal støtte. Desuden viste observationer og kommentarer fra forsøgspersonerne at de havde svært ved at forstå det tematiske kort og fiskeøjegrænsefladen, og at dette sænkede tilfredsheden ved systemerne. Graferne over faser i læseprocessen hjælper også til at forstå forskelle mellem grænsefladerne og forstå hvordan de støtter læsning og navigation i dokumenter.

Sammenfattende er min vigtigste pointe at brugergrænsefladen og interaktionen mellem datamaskiner og mennesker har afgørende indflydelse på informationssøgning. Mit forslag er derfor at fokusere på brugs- og læsevenlighed ved design og evaluering af informationssøgesystemer.

English Summary

Information retrieval systems support searching, accessing, reading, and organizing information. This paper discusses how to evaluate such systems.

From a background in computer science and human-computer interaction, I argue that usability and readability should be key concepts in any evaluation of information retrieval systems. Usability concerns the effectiveness, efficiency, and satisfaction with which users solve tasks. Readability concerns the ease and satisfaction with which users read and comprehend documents and the support for navigation in the documents. These concepts resonate well with current attempts in Information Retrieval to move away from using recall and precision as evaluation measures. In addition, usability and readability emphasize the importance of users interfaces.

The utility of usability and readability in evaluating information retrieval systems is illustrated with two examples. One example presents an evaluation of a thematic map interface. Thematic maps show documents arranged on a map in a way that reflect their thematic similarity. The map also shows words that describe the themes in the document collection. The evaluation compared thematic maps to a text-based information retrieval system on usability measures such as interactive recall, task completion time, and satisfaction. The example shows how different usability measures, such as time and satisfaction, may contradict each other. In the example a detailed analysis reveals how subjects use the map for searching. This analysis shows how the map is misunderstood and that subjects sometimes browse the map aimlessly.

The second example concerns three interfaces that support reading and navigation of scientific papers. Information visualization techniques were used to create an overview+detail interface and a fisheye interface. Using these interfaces, 20 subjects wrote essays and answered questions about scientific documents. We find that subjects prefer the overview+detail interface and with this interface write essays that receive a higher grade. Subjects complete essays faster with the fisheye interface, but seem to gain a less complete understanding of the documents read. In addition, we studied the reading process through visualizations of how subjects read the documents. This analysis reveals that the overview distract the subjects and lead to unnecessary exploration of the document. The visualizations of reading patterns also show how the fisheye interface lead to a more overview-oriented reading style.

In conclusion, we find usability and readability to provide useful notions on which to base evaluations of information retrieval systems. Our examples show that different aspects of usability might contradict each other. The studies of readability suggest that user interfaces may better improve support for reading – which is virtually unexplored in Information Retrieval. In both examples detailed studies of the interaction between users and interfaces leads to plausible hypotheses about the differences in usability and readability. We suggest that usability and readability be systematically explored in future evaluations of information retrieval systems.

References

- Bishop, A. P. 1999. Document structure and digital libraries: how reseachers mobilize information in journal articles. *Information Processing and Management*, 35, 255-279.
- Borg, I. & Groenen, P. 1997. *Multidimensional Scaling*. Springer Verlag.
- Bradow, R., Mitze, K., and Rau, L. F. 1995. Automatic condensation of electronic publications by sentence selection. *Information Processing & Management*, 31, 5, 675-685.
- Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. 1999. The tangled web we wove: a taskonomy of WWW use. I *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '99*, Pittsburgh PA, Maj 1999). ACM Press, New York, NY, 544-551.
- Card, S. K., Mackinlay, J. D. & Shneiderman, B. 1999. *Readings in Information Visualization*. San Francisco CA: Morgan Kaufmann
- Chen, H., Houston, L., Sewell, R., and Schatz, R. 1998. Internet browsing and searching: User evaluations of category map and concept space techniques. *Journal of the American Society of Information Systems*, 49, 7, 582-603.
- Chin, J. P., Diehl, V. A., & Norman, K. L. 1988. Development of an instrument for measuring user satisfaction of the human-computer interface. I *Proceeding of the ACM Conference on Human Factors in Computing Systems (CHI '88*, Washington, DC, Maj 1988). ACM Press, New York, NY, 213-218.
- Dillon, A. 1994. *Designing Usable Electronic Text*. London: Taylor & Francis.
- Dumais, S. T. 1996. Interface Design for Information Retrieval. In *Workshop on Social Aspects of Digital Libraries*. Se <http://is.gseis.ucla.edu/research/dl/dumais.html>.
- Frøkjær, E., Hertzum, M., & Hornbæk, K.. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? I *Proceeding of the ACM Conference on Human Factors in Computing Systems (CHI 2000*, Haag, Holland, Maj 2000). ACM Press, New York, NY, 345-352.
- Hearst, M. A. 1995. TileBars: visualization of term distribution information in full text information access. In . I *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '95*, Denver CO, Maj 1995). ACM Press, New York, NY, 59-66.
- Helms Jørgensen, A. 1990. Thinking-aloud in user interface design: a method promoting cognitive ergonomics. *Ergonomics*, 33, 4, 501-507.
- Hersh, W. 1994. Relevance and Retrieval Evaluation: Perspectives from Medicine. *Journal of the American Society for Information Science*, 45, 3, 201-206.

- Hornbæk, K. & Frøkjær, E. 1999. Do Thematic Maps Improve Information Retrieval? In *IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT '99*, Edinburgh, Scotland, August 1999). IOS Press, Amsterdam, 179-186.
- Hornbæk, K. & Frøkjær, E. 2001. Reading Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2001*, Seattle, WA, Marts 2000). ACM Press, New York, NY, 293-300.
- ISO 1998. Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: Guidance on Usability.
- Kengeri, R., Seals, C. D., Harley, H. D., Reddy, H. P., and Fox, E. A. 1999. Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. *International Journal of Digital Libraries*, 2, 157-169.
- Marchionini, G. 1992. Interfaces for End-User Information Seeking. *Journal of the American Society for Information Science*, 43, 2, 156-163.
- Marchionini, G. and Komlodi, A. 1998. Design of Interfaces for Information Seeking. *Annual Review of Information Science and Technology*, 33, 89-130.
- Rothkopf, E. 1971. Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behaviour*, 10, 608-613.
- Salton, G. & McGill, M. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill.
- Schrivver, K. 1997. *Dynamics in Document Presentation*. New York/London: John Wiley & Sons.
- Sellen, A. and Harper, R. Paper as an analytic resource for the design of new technologies. I *Proceeding of the ACM Conference on Human Factors in Computing Systems (CHI'97*, Atlanta GA, Marts 1997). ACM Press, New York, NY. 319-326,
- Shneiderman, B. 1998. *Designing the User Interface*. 3. udgave. Reading, MA: Addison-Wesley.
- Young, D. and Shneiderman, B. 1993. A Graphical Filter/Flow Representation of Boolean Queries: A Prototype Implementation and Evaluation. *Journal of the American Society of Information Systems*, 44, 6, 327-339.