

Summa – et bud på fremtidens søgesystem

Af Birte Christensen-Dalsgaard

26. oktober blev Statsbibliotekets brugere mødt af et nyt søgesystem, som på mange måder var revolutionerende i forhold til det interface, traditionelle bibliotekssystemer i dag stiller til rådighed

Birte Christensen-Dalsgaard
Udviklingschef
bcd@statsbiblioteket.dk



Fra dagens nyhed:

Fra i dag ser det anderledes ud, når du søger via søgefeltet på forsiden af Statsbiblioteket.dk. Statsbibliotekets søgesystem er helt nyt og giver en række forbedringer, der gør livet en del lettere for brugerne.

Det nye søgesystem medfører en lang række forbedringer. Heriblandt:

- Søgning går hurtigt
- Tilgængelighed af bøger m.v. kan ses helt fremme i søgeresultatet, så du ikke behøver at klikke dig ind på hvert enkelt materiale for at tjekke, om eller hvornår du kan låne det
- Du får sorteret søgeresultatet efter relevans eller udgivelsesår, hvis du ønsker det
- Systemet kommer i mange tilfælde med med stavforslag, hvis du taster forkert (kendt fra Googles "mente du...?")
- Du kan indsnævre dit søgeresultat via links til højre for søgeresultatet (f.eks. begrænse til en bestemt forfatter eller et emne). Forslagene til indsnævring genereres ud fra netop din søgning
- Det er let at bestille flere ting på én gang - læg tingene i materialekurven og bestil tingene på stribe.

Søgesystemet er Statsbibliotekets første bud på, hvordan en søgning i disse Web2.0- og Bibliotek2.0-tider kan se ud og på, hvordan en faglig formidling kan indarbejdes i et biblioteks søgesystem. Vi har hentet ideer fra Google,

Amazon, Flickr, Delicious og har i samarbejde mellem Statsbibliotekets biblioteksfaglige og it-faglige personale lavet et produkt, der kan hjælpe brugeren til at blive virkelig selvhjulpne.

Summa er derfor et projekt, som ud over udvikling af en søgemaskine også omfatter overvejelser over funktionaliteten i fremtidens søgegrænseflade og eksperimenterer med konkrete bud på, hvordan disse ser ud.

Bibliotekets rolle i udvælgelsesfasen

At bibliotekssektoren har et problem med at være synlig i brugerens miljø er ikke en ny erkendelse. Mange undersøgelser [1, 2] har vist, at brugerne foretrækker andre værktøjer. I forbindelse med et feltstudium om brugernes forventninger til det hybride bibliotek [3] blev det yderligere klart, at en meget stor del af brugerne – med eller uden ret betragter sig som selvhjulpne og derfor ikke søger professionel bibliotekarisk hjælp i forbindelse med materialesøgning. Denne brugergruppe udvælger materialet via f.eks. Google, Amazon, anbefalinger fra kolleger eller via referencer i kendt litteratur, og henvender sig først til bibliotekssystemet for bestilling via hjemmesiden, når materialet er fundet. Selve den proces, hvor opdagelsen og udvælgelsen af materiale sker, er altså i mange tilfælde flyttet uden for bibliotekets rammer.

Spørgsmålet er selvfølgelig om bibliotekerne overhovedet har og skal have en rolle i selve materialeudvælgelsesprocessen, hvor brugeren skal tage stilling til spørgsmål som:

- Er denne artikel den mest relevante for min opgave?
- Er det nødvendigt at udforske, om der er mere relevant materiale for at jeg kan skrive min afhandling vederhæftigt?
- Har jeg fundet de centrale værker til brug for det, jeg arbejder med lige nu?

Traditionelt er denne vejledning givet gennem direkte kontakt mellem brugeren og bibliotekspersonalet, idet de kendte bibliotekssystemer kun i meget lav grad understøtter udvælgelsen. Litterære sammenhænge mellem materialer og autoriteters anbefalinger er simpelthen ikke afspejlet i dagens systemer, da de jo primært er konstrueret til bestilling og ikke udvælgelse.

Summa bygger på en tro på, at biblioteket har en rolle i at hjælpe med at finde frem til det bedste materiale. Men også en tro på, at bibliotekerne rent faktisk ligger inde med data og viden, som gør os i stand til at levere et helt unikt produkt. Summa er ikke en konkurrent til Google, men et alternativ bud, som forsøger at forene det bedste fra Google og biblioteksverdenen.

Vi har hentet ideer fra Google, Amazon, Flickr, Delicious og har i samarbejde mellem Statsbibliotekets biblioteksfaglige og it-faglige personale lavet et produkt, der kan hjælpe brugeren til at blive virkelig selvhjulpne.



Teamet bag Summa udviklingen

Stående fra venstre: Jens Hofman Hansen, Kåre Fiedler Christiansen og Hans Lund. Halvsiddende Michael Nielsen.

Siddende bagerst: Bolette Ammitzbøll Madsen, Dorete Larsen og Hans Lauridsen.

Forrest sidder fra venstre Toke Eskildsen, Jørn Thøgersen og Mikkel Erlandsen. Ikke med på billedet er Mads Villadsen.

Foto: Lars R. Clausen, Statsbiblioteket

Syntetisering, specialisering og mobilisering

Summa arbejder mod at understøtte den kvalificerede udvælgelsesproces ved at indarbejde så meget viden om informationer, som muligt og ved at tage en holistisk tilgang til informationsøgning. Det var derfor med stor fornøjelse, at jeg for nylig læste en artikel af Murray [4], som argumenterede for at se på de tre funktioner: "synthesise, specialise og mobilise":

- Synthesise er det at kombinere en diversitet af materiale til et samlet hele
- Specialise er under brug af specifik viden at udvælge og/eller samle materiale, så det kan bruges i specifikke sammenhænge
- Mobilise er at sikre, at materialet kan tilgås på en for brugeren naturlig måde.

Set i forhold til ovenstående er kernen i Summa syntetisering.

Generelt vil **Syntetisering** bestå i at indsamle og udnytte informationer fra en lang række kilder. Det kan være traditionelle informationer som dokumenter, beskrivelse af

dokumenter såsom biblioteksposter og meta-data fra Institutional Repositories, abstracts, indholdsfortegnelser osv., men det kan også være andre typer af informationer. En illustration på brugen af "anderledes" informationer er rangordning på basis af, hvor mange biblioteker, der har en given bog (Open WorldCat). Et andet eksempel er en analyse af, hvilke typer biblioteker bogen står hos, som resulterer i en indikation af bogens målgruppe, "audience level" [5]. Flere er begyndt at bruge udlåns- eller forbrugsstatistik til at rangordne materiale, og endelig er der en vis fokus på aktiv brug af informationer som eksempelvis impactfaktoren. Ideen er, at alle disse informationer kan indgå i en automatisk rangordning af søgeresultatet, og dermed kan den biblioteksfaglige viden være med til at afgøre, hvilke materialer brugeren får tilbudt.

I Summa bruger vi ordet "relevans" til vores rangordning. Men hvad er relevans? Vi ser som udgangspunkt, at relevans er et produkt af tre faktorer: Hvor godt matcher posten en søgning,

hvor god er kvaliteten af det materiale, som repræsenteres af posten og endelig, hvor godt matcher materialet brugerens behov?

I Summa har vi indtil nu arbejdet intensivt med første led: Hvor godt matcher posten søgningen. I grove træk udregner vi en score for en post ved at se på: Hvor vigtig er placeringen af søgetermerne i posten, hvor tæt står søgetermerne sammen i posten og hvor vigtige er søgetermerne? Vigtigheden af en søgeterm kan kun udregnes ved at kende samtlige ord og deres middelhypighed i det totale indeks – hvilket i øvrigt også er den teoretiske grund til, at sam-søgningssystemer i flere samtidige baser aldrig vil kunne levere en fornuftig rangordning af et samlet søgeresultat. Næste iteration af Summa tager fat på problematikken omkring, hvor god posten er. Vi vil her opsætte bibliometriske algoritmer ved at inddrage udlånsstatistikker, referencer mellem materialer, beholdningsoplysninger mv i beregning af relevans.

Den sidste faktor, hvor godt matcher materialet brugeren, hænger sammen med Murray's

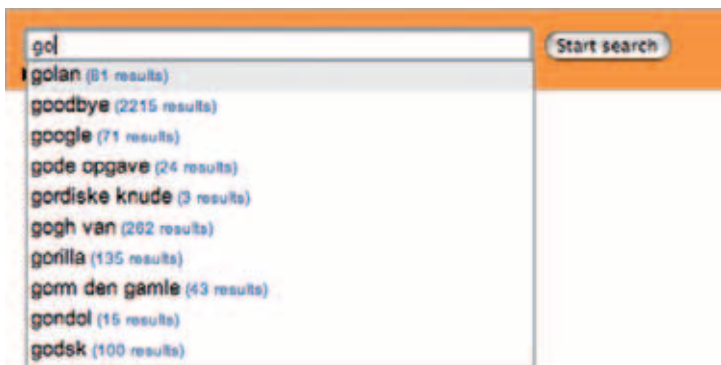
Søgning på: kønsstudier. (Alle materialer)

1 til 3 af 3. Forrige 1 Næste Sortér efter: [År](#) | Relevans

	Titel	Forfatter	År	Type	Kurv
Hjemme	Frem mod år 2000 : kvinde- og kønsforskning - ligestilling	Bente Rosenbeck, Nina Lykke, Foreningen for Kvinde- og Kønsforskning i Danmark. Universitetsudvalget	1992	bog	Tilføj
Online	 Fagekspert i antropologi/etnografi, medicinsk antropologi, etnologi, kulturstudier, religionshistorie og kønsstudier	Eva Lous		person	Tilføj
Special	Årsberetning ved generalforsamlingen ... / Foreningen for Kvinde- og Kønsforskning i Danmark			tidsskrift	Tilføj

1 til 3 af 3. Forrige 1 Næste Sortér efter: [År](#) | Relevans

Figur 1: Vi har eksperimenteret med at lægge beskrivelser af fagekspert ind i datagrundlaget for Summa. På denne måde bliver den faglige ekspertise eksponeret som en integreret del af resultatet af en søgning.



Figur 2: Når man begynder at indtaste sit søgeord, vil Summa løbende komme med forslag, som er listet efter, hvor mange gange de har været brugt i søgninger, som resulterede i resultat med hits.

specialisering. Det er her, at eksempelvis en artikels alder gennemgående har stor betydning indenfor naturvidenskab, eller man kun er interesseret i specielle typer af informationer, kan indarbejdes i rangordningen. Specialisering kan enten ske gennem specielle indgange for bruger kategorier eller bedre ved at introducere personlige profiler eller metoder til automatisk at udnytte tidligere adfærd eller mulighed for at træne søgesystemet – på samme måde, som man kan træne eksempelvis mailfilteret i mange mailprogrammer.

I første omgang er mobiliseringen simpel – et enkelt søgefelt, som til gengæld kan placeres rigtig mange steder, hvor det findes relevant. Dette er vist på figur 2, hvor vi også

demonstrerer at systemet understøtter en forslagsfunktion¹, inspireret af Google Suggest. Indtaster man et ord, som systemet ikke kender, vil det komme med forslag til, hvad der kunne have stået, ”Mente du”, som også kendes fra Google.

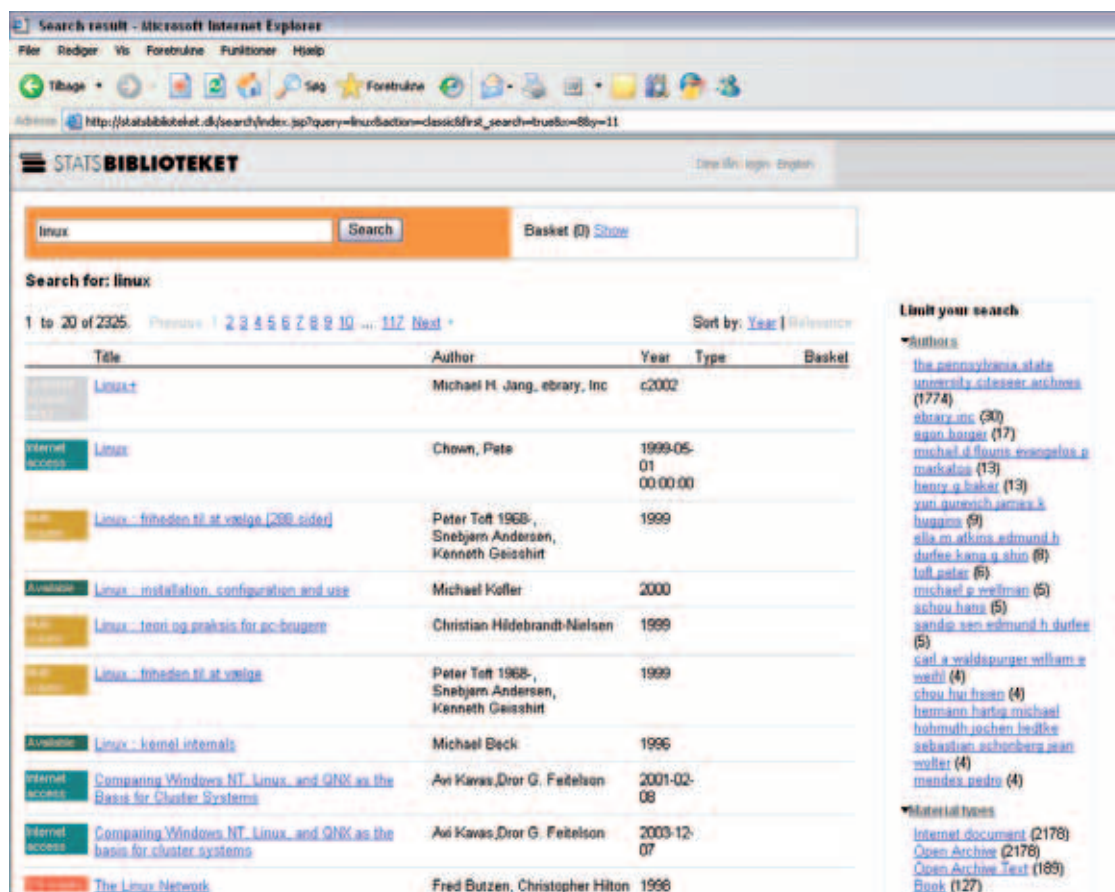
Præsentationssystemet

Summa baserer sig på nye præsentationstekniker og mere åbne standarder end traditionelle bibliotekssystemer. Det giver dels frihed til at vælge blandt alle de metoder og ideer, der findes blandt internetudviklere, dels stiller det brugeren i den situation, at bibliotekets grænseflade og søgeværktøjer mere ligner det, han er vant til at bruge.

Det har været et mål for udviklingen af Summa at få adskilt søge- og præsentationssystemet fra bibliotekssystemet, således at vekselvirkningen mellem de to er velbeskrevet og sker gennem brug af webservices. Det gør, at Summa kan ligge oven på en vilkårlig mængde data kilder og muliggøre søgning direkte i alle data på én gang. Til forskel fra sædvanlige samsøgningsystemer kan Summa præsentere ét samlet søgesæt, der kan sorteres efter relevans eller andre kriterier, som brugeren måtte finde nyttige i sin stræben efter at finde netop det rigtige materiale.

Summa er ikke et bibliotekssystem, men et system, som er designet til at arbejde sammen med et eller flere bestående bibliotekssystemer, samtidig med at materiale fra mange andre kilder kan inkluderes. Summa kan således ikke erstatte et bibliotekssystem, men kan erstatte de ofte ret rudimentære søgemekanismer, som bibliotekssystemerne er født med.

I udviklingen af søgegrænsefladen blev der eksperimenteret med flere forskellige visninger af søgeresultaterne, og denne eksperimenteren er langt fra færdig. Oversigtsbilledet med resultatet af en søgning har indbygget materialestatus og bestilling, som hentes ind asynkront, idet vi har benyttet AJAX² i udviklingen af præsentationssystemet.



Figur 3: Søgeresultat. Statusinformation hentes via en webservice. Indsnævringsmulighederne i højre side er resultatet af en realtidsanalyse af søgeresultatet.

Til forskel fra sædvanlige samsøgningssystemer kan Summa præsentere ét samlet søgesæt, der kan sorteres efter relevans eller andre kriterier, som brugeren måtte finde nyttige i sin stræben efter at finde netop det rigtige materiale.

Posterne fra en søgning analyseres i realtid for information om forfatter, materialetype, opstilling osv. og resultatet af denne analyse vises i højre side af oversigtsbilledet. Denne liste giver mulighed for at afgrænse søgeresultatet til eksempelvis de bøger, som står på et institutbibliotek³. (Se figur 3)

Når en post vises, beriges visningen med information fra en række webservices. Ideen er at give de ekstra informationer, som kan hjælpe brugeren til at afgøre, om den fremfundne post repræsenterer det materiale, han/hun søger. Arbejdet bygger på erfaringer, som er opnået i forbindelse med "Mit bibliotek" projektet [6]. Vi fletter indhold ind fra Forfatterweb (forfatterportrætter), Litteratursiden (anbefalinger og forfatterportrætter), Bibliotekernes Netmusik (cd-covers og musikprøver), Amazon (bogforsider og anmeldelser) og Mit bibliotek ("andre har lånt"-funktion) (se figur 4).

Grænsefladen stiller desuden en opsamlingskurv til rådighed for brugeren, sådan at opsamlede referencer kan udskrives, bestilles, sendes som e-mail og eksporteres som referencer til referenceværktøjer som RefWorks.

Der tilbydes standard fuldtekstsøgning samt en feltbaseret søgning i de vigtigste felter (forfatter, titel, årstal). Grænsefladen understøtter kommandobaseret søgning i fuldtekst-søgefeltet, hvorved bibliotekarer i en og samme grænseflade som slutbrugerne får et meget præcist værktøj til specialsøgninger.

Datagrundlag

Målet er, at Summa skal indekseres alt kvalitetsmateriale, som er tilgængeligt for forskere og studerende ved Aarhus Universitet.

En af udfordringerne i forbindelse med et fælles datagrundlag er forskelligheden i formen, mængden og kvaliteten af data. Vi har alt fra meget strukturerede MARC-poster til temmelig sparsomme Dublin Core poster, og alt fra en simpel metadatabeskrivelse til en fuld, struktureret tekst. Det er klart, at det er en udfordring at lave en relevant rangordning på basis af sådanne uhomogene data. Vi har besluttet ikke at smide data væk, men at benytte en metode, hvor en evt. struktureret del kan gøres sammenlignelig og informationsindholdet i resten udnyttes. Vi har derfor fravalgt at konvertere data til samme format. I stedet konverteres alle data til XML, men beholdes i øvrigt deres formater og fulde indhold. Sammenlignelighed opnås gennem en vægtet formatering, hvor der skal udvikles vægte for hver datakilde. Ved hjælp af XSL-stylesheets laves indeksering, rangordning, kort format og fremvisningsformat for hver enkelt datatype.

Vores datagrundlag er ved lanceringen af søgefladen sammensat af:

- De hidtidige data fra vores Horizon bibliotekskatalog – dvs:
 - bibliografiske data og beholdningsdata for materialer på Statsbiblioteket og institutbibliotekerne på Aarhus Universitet

- bibliografiske data og beholdningsdata fra enkelte andre forskningsbiblioteker i Århus-området, hvor vi yder biblioteksteknisk service (f.eks. Arkitektskolen i Århus)
- bibliografiske data fra Århus Universitets-hospital
- bibliografiske data for e-bøger
- En beskrivelse af vores fagreferencer, som derfor kan indgå i søgeresultaterne
- Metadata fra registrering af digitaliserede, historiske reklamefilm
- OAI-metadata høstes i øjeblikket fra en stribe udvalgte targets⁴

Hvad angår bibliografiske data fra de naturvidenskabelige institutbiblioteker ved Aarhus Universitet, som har et Aleph bibliotekssystem, har vi lavet en aftale om, at vi får dataleverancer med deres poster. Men da der endnu ikke er udviklet de nødvendige webservices over for Aleph systemet, kan dette materiale ikke bestilles og status kan ikke ses.

Det svage punkt er registreringer på artikelniveau. Et første skridt mod at få dem inkluderet er taget gennem en aftale med Cambridge Scientific Abstracts (CSA) om, at vi kan få deres data og forsøgsvis lægge dem ind i søgefladen. Ved fremtidige licensaftaler håber vi at kunne ændre denne situation, således at vi får mulighed for at få leveret metadata (og meget gerne fuldtekst til indekseringsformål) sammen med de elektroniske adgange.

[Tilbage til søgeresultat](#)

Små blå breve (musik/lyd)

Forfatter/ophav	Nanna, Kim Sagild, Søs Fenger, Peter Biker, Tamra Rosanes, Helge Solberg, Martin Taarup, Martin Nordborg, Jacob Andersen, Klaus Menzer, Inge Mulvad Eje, Jens Haack 1960-08-19-
Titel	Små blå breve / teknik: Jørgen Knub, Mads Nilsson, Anders Valbro og Werner Scherrer / mix: Jørgen Knub / producer: Kim Sagild
Sted	Replay, P 1982, C 1998
Status	Udsolgt

[Bestil](#)

[Flere oplysninger](#)

Cover fra [Netmusikken](#)



Tracks

[Helst om kærlighed](#)
[Deja vu](#)
[Små, blå breve](#)
[Rolling stone](#)
[Buster](#)
[Havet er livligt i dag](#)
[Sammenhold, du](#)
[Anytime at all](#)
[10/4 Nyhavn](#)
[Tankespind](#)

Figur 4: Resultat af søgning på en CD, hvor musikken også er til rådighed via Bibliotekernes Netmusik; cover vises og der gives mulighed for at høre de første 30 sek. af de tilgængelige tracks.

Selv om Summa kan siges at tage et kvantespring frem, er der stadig meget, der skal laves og mange tanker, som skal tænkes og afprøves. Det er derfor meget vigtigt for Statsbiblioteket at have rådighed over en åben platform, hvor der kan eksperimenteres med, hvordan biblioteks-faglig formidling kan ske.

Det videre forløb

Arbejdet med brugergrænsefladen er stadig i gang, og den første brugbarhedstest er gennemført. Fem brugere gennemførte en tænkehøjt-test. Det er klart, at der stadig er meget at lave og tilpasse, men overordnet set var reaktionerne meget positive, og grænsefladen blev rost for:

- Stort overblik i søgeresultatet
- Godt overblik i fuld post
- Nyttigt at udlånsstatus er så tydelig
- Nyttig kurv-funktionalitet

Summa er opbygget som et meget modulært og skalerbart produkt. Det kan omfatte mange indekser, mange maskiner, mange datakilder, mange data og mange samtidige brugere. Summa kan konfigureres og tilpasses på mange niveauer, således at mange forskellige brugsscenarier kan serviceres fra det samme grundsystem.

Summa er allerede brugt som basis for Bibliotekernes Netmusik, og i det kommende år vil der blive arbejdet på to fronter omkring udvikling af formidlingssystemet, dels i relation til bred tværinstitutionel kulturformidling og dels i relation til bred biblioteksformidling. Det sidste vil ske i samarbejde med en række folke-, central- og forskningsbiblioteker⁵. Silkeborg Bibliotek, Århus Kommunes Biblioteker, Københavns Kommunes Biblioteker, Odense Centralbibliotek, Frederiksberg Kommunes Biblioteker, Gentofte Bibliotekerne, Vejle Bibliotek og Herning Centralbibliotek deltager aktivt i testning og udvikling af præsentationssystemet. Samtidig er der dannet en følgegruppe, som består ud udviklingsbibliotekerne, Biblioteket ved Handelshøjskolen i Århus, Roskilde Universitetsbibliotek samt Randers Bibliotek, som vil blive inddraget i den løbende dialog omkring biblioteksfaglig relevant formidling af søgeresultater. Følgegruppen kan blive udvidet, da vi stadig er i dialog med potentielle deltagere. Da de fleste af de deltagende biblioteker har et Axiell bibliotekssystem, er der også indledt en dialog med dette firma.

Selv om Summa kan siges at tage et kvantespring frem, er der stadig meget, der skal laves og mange tanker, som skal tænkes og afprøves. Det er derfor meget vigtigt for Statsbiblioteket at have rådighed over en åben platform, hvor der kan eksperimenteres med, hvordan biblioteks-faglig formidling kan ske.

Tak

Tak til alle, som har bidraget til denne artikel. Jeg overvejede en overgang at sætte hele Summa udviklingsgruppen på som forfattere, men fandt at en forfatterliste – alfabetisk efter fornavn: Bolette Ammitzbøll Madsen, Dorete Larsen, Hans Lauridsen, Hans Lund, Jens Hofman Hansen, Jørn Thøgersen, Kåre Fiedler Christiansen, Mads Villadsen, Michael Poltorak Nielsen, Mikkel Erlandsen og Toke Eskildsen, nok ville være for omfattende.

Referencer

[1] Niels Ole Pors, 2005, "Studerende, Google og biblioteket: En undersøgelse af 1694 studerendes brug af biblioteket og informationsressourcer". Set d. 16.11.2006 på: www.bs.dk/publikationer/andre/studerende/index.htm

[2] OCLC Membership report, 2006. *College Students' perceptions of Libraries and Information Resources*. Set d. 16.11.2006 på: www.oclc.org/reports/perceptionscollege.htm

[3] Jeppe Lomholt Akselbo, Lise Arnfred, Sten Barfort, Gina Bay, Tine Bagger Christiansen, Jens Hofman Hansen, Henrik Tvermoes Jensen, Gitte Bach Markussen, Anna Mette Morthorst og Michael Poltorak Nielsen, 2006. *Det hybride bibliotek set med brugernes øjne*. Set d. 16.11.2006 på: www.statsbiblioteket.dk/publ/feltstudier.pdf

[4] Robin Murray, 2006. *Library Systems: Synthesis, Specialise, Mobilise*. Set d. 16.11.2006 på: www.ariadne.ac.uk/issue48/murray/

[5] Lavoie, Brian, Lorcan Dempsey, and Lynn Silipigni Connaway. 2006. *Making Data Work Harder*. *LibraryJournal.com*, January 15, 2006. Set d. 16.11.2006 på: www.libraryjournal.com/article/CA6298444.html

[6] *Mit bibliotek*. Udviklingsprojekt under ledelse af Århus Kommunes Biblioteker og udført i samarbejde med Frederiksberg Kommunes Biblioteker, Gentofte Bibliotekerne, Herning Centralbibliotek, Københavns Kommunes Biblioteker, Silkeborg Bibliotek, Vejle Bibliotek, Statsbiblioteket og Dansk BiblioteksCenter

Noter

¹ Summa – suggest.: Alle søgninger i Summa registreres automatisk for senere at kunne give brugerne forslag til søgninger, mens de skriver deres søgning. I Summas suggest funktion vægtes forslag efter, hvor populær en søgning er hos brugerne, således at den oftest udførte søgning, der matcher det brugerne har tastet, vises som det første forslag. Funktionen fungerer ved, at der ved hvert tastetryk fra brugeren sendes en forespørgsel til suggest modulet, der sender en liste med forslag tilbage til brugerens browser.

² AJAX (Asynkron JavaScript og XML) er en teknologi, som populært fortalt tillader udvikling af websider, der opfører sig som en klient ved løbende at kunne reagere på, hvad der sker, og som giver mulighed for en gradvis opbygning af websiden. Fordelen er, at brugeren umiddelbart ser resultaterne, og at disse gradvis udbygges, efterhånden som de forskellige services svarer.

I Summa har vi været pragmatiske i vores brug af AJAX-teknologi i den forstand, at den hovedsagelig anvendes til at præsentere brugeren for ekstratilbud. Det vil sige, at præsentationen af det basale søgeresultat og den basale postvisning foregår på Web 1.0-manér - alt genereres på serveren og skubbes så ud til brugeren. Når brugeren så har fået søgeresultatet/postvisningen, går vi i gang med at berige det vha. AJAX. Det drejer sig fx om at hente status dynamisk ind på siden, vise forsider fra Bibliotekernes Netmusik og Amazon, "Andre, der har lånt", Forfatterweb, mm. Herudover benyttes AJAX til forslagsfunktionen og til at hente og gemme oplysninger om de poster, brugeren putter i kurven.

³ Effektive metoder til at danne indekser og til at efterbehandle søgeresultatet i realtid har været en vigtig aktivitet. Eksempelvis kan nævnes, at systemet er udviklet med henblik på at håndtere op imod 100 millioner poster. Alternativt tænkning var derfor nødvendig, og løsningen blev etableringen af et GRID bestående af de pc'er, som medarbejderne bruger på Statsbiblioteket. Vi benytter samme metode som Google, Mapreduce. Indeksering sker nu på Statsbibliotekets egen distribuerede platform bestående af kontor PC'er og er i dag optimeret til at udnytte 10-50 maskiner. Platformen er meget generisk og kan benyttes til andre formål end blot indeksering.

⁴ Data fra OAI (Open Archive Initiatives) OAI-baserne er udvalgt, så de dels repræsenterer forskellige datatyper (publikationer, digitale objekter og datasæt), der muliggør, at vi kan eksperimentere med forskellige måder, hvorpå forskellige datatyper præsenteres, men udvalgt således, at vi får erfaring med at håndtere dataopsamling fra nogle af de store OAI-udbydere. Vi anvender i øjeblikket data fra: Bioline International, Directory of Open Access Journals, Library of Congress Open Archive, Scientific Literature Digital Library (CiteSeer), PubMed Central, HighWire Press, Citebase, CERN Document Server, Pangaea - Publishing Network for Geoscientific & Environmental Data, NLA - National Library of Australia, Citidel - Computing and Information Technology Interactive Digital Educational Library, BioMed Central, African Journals OnLine, Caltech Library System Papers and Publications, Institute of Physics.

⁵ En generel invitation blev sendt til alle i IT12.